FOR OFFICIAL USE ONLY
ACCESS DB # 95108

PLEASE PRINT CLEARLY
Location (Bldg/Room#): _____

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner # : _____
Art Unit: _____ Phone (30__) _____ Serial Number: _____
Results Format Preferred (circle): PAPER DISK E-MAIL

****

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Date: _____

Search Topic:
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

*"For Sequence Searches Only" ***Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.***

FOR OFFICIAL USE ONLY

****

STAFF USE ONLY

Searcher: HART
Searcher Phone #: 305-9203
Searcher Location: 12C14
Date Searcher Picked Up: 2/9/00
Date Completed: _____
Searcher Prep & Review Time: _____
Online Time: _____

Type of Search

2 NA Sequence (#)
3 AA Sequence (#)
_____ Structure (#)
_____ Bibliographic
_____ Litigation
_____ Fulltext
_____ Other

Vendors and Cost

_____ STN _____ Dialog
_____ Questel/Orbit _____ Dr.Link
_____ Lexis/Nexis _____ Westlaw
_____ WWW/Internet
X In-house sequence systems (list)
_____ Other (specify)

ABSS 02

25108

STIC-Biotech/ChemLib

From: Duffy, Patricia
Sent: Tuesday, February 08, 2000 1:51 PM
To: STIC-Biotech/ChemLib
Subject: Sequence Search 09/060,609

In RE: 09/060,609

Please search SEQ ID NO:1 and a oligomer search of SEQ ID NO:1.
Please search SEQ ID NO:2 and an oligomer search of SEQ ID NO:2.
Please backtranslate the amino acid sequence of SEQ ID NO:2 into NA and run NA.

Please include an interference search.

Thanks,

Patricia A. Duffy
Art Unit 1645; CM-1 8D05
305-7555

```
OM of: US-09-060-609-2 to: GenEmbl:*   out_format : pfs

Date: Feb 10, 2000  6:57 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool/US09060609/runat_09022000_150438_25179/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09060609
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

[illegible]rch information block:
[illegible]ry: US-09-060-609-2
Query length: 269
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 497.180000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_in2:AC007175        -   306.50  431.90  4.9e-16   188633  ! AC007175 Drosophila melanoga
em_htg1:AC004434       +   306.50  439.86  1.8e-16    83204  ! Ac004434 Drosophila melanogas
em_htg1:AC006092       +   306.50  435.33  3.2e-16   132637  ! Ac006092 Drosophila melanoga
gb_htg2:AC005718       -   257.00  354.51  1.0e-11   149592  ! AC005718 Drosophila melanoga
gb_htg1:DMBR30C13      +   183.50  238.59  2.9e-05   117674  ! AL122025 Drosophila melanoga
gb_in2:AF181623        +   183.50  281.41  1.2e-07     1440  ! AF181623 Drosophila melanogast
gb_in2:CELC41D11       -   162.00  214.27  0.0007     40912  ! AF003740 Caenorhabditis elega
gb_htg6:AC016672       +   139.00  161.59  0.5613    205157  ! AC016672 Homo sapiens clone
gb_in1:CELC02F5        +   138.50  182.35  0.0391     22333  ! L14745 C. elegans cosmid C02F
gb_htg6:AC011069       +   132.00  152.25  1.86      168266  ! AC011069 Drosophila melanoga
gb_ba2:AE001582        -   117.00  149.46  2.66       18753  ! AE001582 Borrelia burgdorferi
gb_pr2:HSAC001228      -   111.00  114.84  225.42     244254  ! AC001228 244Kb Contig from H
gb_pr3:AC005950        -   111.00  120.40  110.52     137932  ! AC005950 Homo sapiens Chroms
gb_htg7:AC013791       +   111.00  117.80  154.13     180077  ! AC013791 Homo sapiens clone
gb_htg3:AC010844       +   110.50  123.48  74.45       92509  ! AC010844 Drosophila melanogas
gb_pr3:AF079117        +   108.00  159.55  0.7286      1498  ! AF079117 Homo sapiens equilibr
[illegible]_pr3:HSU81375   +   108.00  155.98  1.15        2162  ! U81375 Human placental equilib
[illegible]_ba1:MTY13E10   -   105.50  124.88  62.20       35019  ! Z95324 Mycobacterium tubercul
[illegible]_ba2:AE001584   +   105.00  120.05  115.55      52971  ! AE001584 Borrelia burgdorferi
gb_ba1:D90844          -   103.50  130.64  29.73       13918  ! D90844 E.coli genomic DNA, Kc
gb_ba1:D90845          -   103.50  127.42  44.89       19366  ! D90845 E.coli genomic DNA, Kc
gb_ba1:ECOALKA         +   103.50  149.73  2.57         1954  ! K02498 E. coli alkA gene encod
gb_vi:EHVU20824        -   102.00  103.09  1.0e+03    184427  ! U20824 Equine herpesvirus 2,
gb_htg1:AP000807       -   102.00  102.42  1.1e+03    197580  ! AP000807 Homo sapiens chromo
gb_htg3:AC008594       +   102.00  120.61  107.58      30450  ! AC008594 Homo sapiens chromos
gb_pr4:AC007038        +   101.00  107.53  575.65      99030  ! AC007038 Homo sapiens clone N
gb_pl2:GHU37060        -   100.50  149.33  2.70         1239  ! U37060 Gossypium hirsutum asc
gb_ba2:AE000297        -    98.50  119.83  118.94      18495  ! AE000297 Escherichia coli K-1
gb_htg2:AC005020       -    98.50   97.82  2.0e+03    177756  ! AC005020 Homo sapiens clone
gb_pr3:HSL113B6        -    98.00  111.58  342.49      39756  ! Z69837 Human DNA sequence fro
gb_pr4:AC005516        -    98.00   95.85  2.6e+03    200348  ! AC005516 Homo sapiens Chromo
gb_htg6:AC012520       +    98.00   98.70  1.8e+03    149500  ! AC012520 Homo sapiens clone
gb_ba1:TAU52917        +    97.00  119.52  123.64      14886  ! U52917 Thermus aquaticus ther
gb_ba2:AE002064        +    97.00  120.32  111.63      13715  ! AE002064 Deinococcus radiodur
gb_pr2:AC002293        +    97.00  117.15  167.62      19000  ! AC002293 Genomic sequence fro
gb_pr3:AC002102        +    97.00  108.97  478.82      44078  ! AC002102 Homo sapiens chromos
gb_pr4:AC004923        +    97.00  101.51  1.2e+03     94851  ! AC004923 Homo sapiens PAC clo
gb_htg6:AF205589       -    97.00   98.30  1.9e+03    131973  ! AF205589 Homo sapiens chromo
gb_pr2:AC001643        -    96.50  109.06  473.30      40203  ! AC001643 Genomic sequence fro
gb_pr3:HSJ392M17       -    96.50   99.19  1.7e+03    110900  ! AL049843 Human DNA sequence
gb_htg2:AC007439       -    96.50   93.38  3.5e+03    201576  ! AC007439 Mus musculus, *** S
gb_pr4:AC005923        -    96.50  101.40  1.3e+03     88326  ! AC005923 Homo sapiens 3p21.3-
gb_htg6:AC005903       -    96.50   91.77  4.3e+03    237907  ! AC005903 Homo sapiens clo
gb_ba2:AF079317        -    95.50   92.63  3.9e+03    184457  ! AF079317 Sphingomonas aro
gb_pl1:PPU08841        +    95.50  137.22  12.77        1882  ! U08841 Porphyra purpurea el

seq_name: gb_in2:AC007175

seq_documentation_block:
LOCUS       AC007175   188633 bp    DNA              INV       27-MAR-1999
DEFINITION  Drosophila melanogaster, chromosome 2R, region 57D3-57E2, BAC
            clones DS01261 and BACR48K03, complete sequence.
ACCESSION   AC007175 AC004434 AC003838 AC003531 AC003532 AC003533 AC006092
VERSION     AC007175.1  GI:4529971
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 188633)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2R, region 57D3-57E2
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 188633)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1999) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     On Mar 27, 1999 this sequence version replaced gi:4510440
            gi:4454570.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library location: 14-13.
FEATURES             Location/Qualifiers
     source          1. .188633
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="57D3-57E2"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
                     /clone="BACs DS01261 (D191) and BACR48K03 (D487)"
                     /note="DS01261 (D191) and R48K03 (D487) were sequenced as
                     a project.  R48K03 extends from BAC end at bp 1 to BAC end
                     at bp 132,637.  DS01261 extends from P1 end at bp 126,093
                     to a minimal overlap with its distal neighbor DS08012
                     (D224) at bp 188,633.  "
BASE COUNT    53352 a  41815 c  41597 g  51869 t
ORIGIN
```

```
alignment_scores:
           Quality:  306.50            Length:    184
             Ratio:  2.620               Gaps:      7
 Percent Similarity:  63.587   Percent Identity: 38.043

alignment_block:
 US-09-060-609-2 x AC007175/rev   ..

 Align seg 1/1  to reverse of: AC007175  from: 1  to: 188633

     107 SerLeuLysCysGluAspLeuLys...ValGlyGlnTyrIleCysLysAs 122
         ::::::   |||:::::::|||:::   :::||||||::::::|||   ||
  127457 AATGTGGACTGCAACGAGCTCCAGATGATGGGCCAGTTCATGTGCCCCGA 127408

     122 pPro.........LysIleAsnAspAlaThrGlnGluProValAsnCysT 136
         ||||         :::|||:::      |||||||:::   ::::::|||| 
  127407 CCCGGCGAGGGGTCAAATCGATCCCAAGACACAGCAGCTGGCCGGATGCA 127358

     136 hrAsnTyr...ThrAlaHisValSerCysPheProAlaProAsnIleThr 151
         ||:::         |||:::|||   |||:::   |||   :::|||::: 
  127357 CCAGGGAGGGTCGCGCCCGCGTCTGGTGCATCGCCGCCAACGAGATCAAC 127308

     152 CysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGl 168
         |||   :::   :::||||||   |||
  127307 TGCACGGAG...ACTGGCAATGCCACG..................... 127284

     168 yPhePheLysProIleSerCysArg..AsnVal.............. 178
          |||   :::   :::   |||:::  :::|||
  127283 .TTCACCCGAGAAGTGCCCTGCAAGTGGACGTGAGTGCAGCTACAGTATC 127235

     179 .....................................AsnG 180
                                              ||||
  127234 TCATAAACATAATCTTAAGACCTATATTCTATTCCTATTCTATAGGAATG 127185

     180 lyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeu 196
         |||||         :::::::::   ||||||:::|||||||||   :::
  127184 GCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTC 127135

     197 GlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPh 213
         |||:::|||||||||||||||||||||||:::::|||||||||||||||
  127134 GGCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTT 127085

     213 eCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuI 230
         |||||||:::|||         :::|||:::|||||||||:::::::|||| 
  127084 CTGCACCCTCGGCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGA 127035

     230 leSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIleAsp 246
         ||:::::::|||:::|||||||||:::|||||||||:::|||:::|||
  127034 TAGCCCTGCAGGTTGTGGGTCCGGCGGATGGCTCCGCCTATGTGATACCC 126985

     247 TyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrPheAr 263
         ||||||||||:::   :::      :::          |||   |||:::||
  126984 TACTACGGAGCCGGCATCCACATCGTGCGCAGCGACAACACCACGTACCG 126935

     263 g 263
         |
  126934 G 126934

seq_name: em_htg1:AC004434

seq_documentation_block:
ID   AC004434   standard; DNA; HTG; 83204 BP.
XX
AC   AC004434; AC003531; AC003532; AC003533; AC003838;
XX
SV   AC004434.4
XX
DT   18-MAR-1998 (Rel. 55, Created)
DT   25-MAR-1999 (Rel. 59, Last updated, Version 17)
XX
DE   Drosophila melanogaster; Chromosome 2R; Region 57D11-57E2; P1 clone
DE   DS01261, WORKING DRAFT SEQUENCE, 10 unordered pieces.
XX
KW   HTG; HTGS_PHASE1.
XX
OS   Drosophila melanogaster (fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC   Drosophilidae; Drosophila.
XX
RN   [1]
RP   1-83204
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster";
RL   Unpublished.
XX
RN   [2]
RP   1-83204
RA   Celniker S.E., George R.A., Galle R., Svirskas R.R., Hoskins R.A.,
RA   Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Chavez C., Chew M.,
RA   Doyle C.M., Farfan D.E., Flanagan J., Houston K.A., Hummasti S.R.,
RA   Karra K., Kearney L., Kim S.H., Lee B., Lomotan M.A., Mak J., Mazda P.,
RA   Mok M.S., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S.,
RA   Pfeiffer B., Punch E., Snir E., Twomey B., Wan K.H., Whitelaw K.R., Yee A.,
RA   Zhang R., Zieran L.L., Kimmel B.E.;
RT   ;
RL   Submitted (17-MAR-1998) to the EMBL/GenBank/DDBJ databases.
RL   Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,
RL   Berkeley, CA 94720, USA
XX
CC   For further information about this sequence, including its location
CC   and relationship to other sequences, please visit our sequence
CC   archive Web site (http://www.fruitfly.org/sequence/) or send email
CC   to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
CC   the following cutoffs: length >= 400 bases, phrap computed error
CC   rate <= 1/10. P1 library location: 14-13.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 10 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1      738: contig of 738 bp in length
CC   *      739      818: gap of unknown length
CC   *      819     1918: contig of 1100 bp in length
CC   *     1919     1998: gap of unknown length
CC   *     1999     2644: contig of 646 bp in length
CC   *     2645     2724: gap of unknown length
CC   *     2725     3214: contig of 490 bp in length
CC   *     3215     3294: gap of unknown length
CC   *     3295     4442: contig of 1148 bp in length
CC   *     4443     4522: gap of unknown length
CC   *     4523     5169: contig of 647 bp in length
CC   *     5170     5249: gap of unknown length
CC   *     5250     6692: contig of 1443 bp in length
CC   *     6693     6772: gap of unknown length
CC   *     6773     9720: contig of 2948 bp in length
CC   *     9721     9800: gap of unknown length
CC   *     9801    13304: contig of 3504 bp in length
CC   *    13305    13384: gap of unknown length
CC   *    13385    83204: contig of 69820 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .83204
```

```
FT                   /chromosome="2"
FT                   /db_xref="taxon:7227"
FT                   /organism="Drosophila melanogaster"
FT                   /strain="y2; cn bw sp"
FT                   /clone="DS01261 (D191)"
FT                   /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
FT                   /map="57D11-57E2"
XX
SQ   Sequence 83204 BP; 23971 A; 17293 C; 17121 G; 24097 T; 722 other;


alignment_scores:
           Quality:  306.50            Length:    184
             Ratio:  2.620              Gaps:      7
 Percent Similarity:  63.587   Percent Identity: 38.043

alignment_block:
 US-09-060-609-2 x AC004434    ..

Align seg 1/1  to: AC004434  from: 1  to: 83204

    107 SerLeuLysCysGluAspLeuLys...ValGlyGlnTyrIleCysLysAs 122
        ::::::   |||::::::|||:::   :::||||||::::::|||   ||
  81846 AATGTGGACTGCAACGAGCTCCAGATGATGGGCCAGTTCATGTGCCCCGA 81895

    122 pPro.........LysIleAsnAspAlaThrGlnGluProValAsnCysT 136
        ||||         :::|||:::      |||||||:::   ::::::|||| 
  81896 CCCGGCGAGGGGTCAAATCGATCCCAAGACACAGCAGCTGGCCGGATGCA 81945

    136 hrAsnTyr...ThrAlaHisValSerCysPheProAlaProAsnIleThr 151
        ||:::        |||:::|||   |||:::   |||   :::|||::: 
  81946 CCAGGGAGGGTCGCGCCCGCGTCTGGTGCATCGCCGCCAACGAGATCAAC 81995

    152 CysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGl 168
        |||   :::   :::|||||||   |||
  81996 TGCACGGAG...ACTGGCAATGCCACG..................... 82019

    168 yPhePheLysProIleSerCysArg..AsnVal................. 178
         |||   :::   :::   |||:::  :::|||
  82020 .TTCACCCGAGAAGTGCCCTGCAAGTGGACGTGAGTGCAGCTACAGTATC 82068

    179 .............................................AsnG 180
                                                    ||||
  82069 TCATAAACATAATCTTAAGACCTATATTCTATTCCTATTCTATAGGAATG 82118

    180 lyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeu 196
        |||||         :::::::::   ||||||:::|||||||||   :::
  82119 GCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTC 82168

    197 GlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPh 213
        |||:::|||||||||||||||||||||||:::::|||||||||||||||
  82169 GGCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTT 82218

    213 eCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuI 230
        |||||||:::|||         :::|||:::|||||||||:::::||||
  82219 CTGCACCCTCGGCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGA 82268

    230 leSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIleAsp 246
        ||:::::::|||:::|||||||||:::|||||||||:::|||:::|||
  82269 TAGCCCTGCAGGTTGTGGGTCCGGCGGATGGCTCCGCCTATGTGATACCC 82318

    247 TyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrPheAr 263
        |||||||||:::   :::      :::        |||   |||:::||
  82319 TACTACGGAGCCGGCATCCACATCGTGCGCAGCGACAACACCACGTACCG 82368

    263 g 263
        |
  82369 G 82369

seq_name: em_htg1:AC006092

seq_documentation_block:
ID   AC006092   standard; DNA; HTG; 132637 BP.

XX
AC   AC006092;
XX
SV   AC006092.3
XX
DT   07-DEC-1998 (Rel. 57, Created)
DT   29-MAR-1999 (Rel. 59, Last updated, Version 10)
XX
DE   Drosophila melanogaster; Chromosome 2R; Region 57D3-57D12; BAC
DE   clone BACR48K03, WORKING DRAFT SEQUENCE, 1 unordered pieces.
XX
KW   HTG; HTGS_PHASE1.
XX
OS   Drosophila melanogaster (fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC   Drosophilidae; Drosophila.
XX
RN   [1]
RP   1-132637
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M♦A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster";
RL   Unpublished.
XX
RN   [2]
RP   1-132637
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Punch E.,
RA   Sequeira A., Sethi H., Snir E., Svirskas R.R., Twomey B., Wan K.H.,
RA   Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;
RT   ;
RL   Submitted (04-DEC-1998) to the EMBL/GenBank/DDBJ databases.
RL   Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,
RL   Berkeley, CA 94720, USA
XX
CC   On Mar 20, 1999 this sequence version replaced gi:4417233.
CC   For further information about this sequence, including its location
CC   and relationship to other sequences, please visit our sequence
CC   archive Web site (http://www.fruitfly.org/sequence/) or send email
CC   to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
CC   the following cutoffs: length >= 400 bases, phrap computed error
CC   rate <= 1/10.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 1 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1   132637: contig of 132637 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .132637
FT                   /chromosome="2"
FT                   /db_xref="taxon:7227"
FT                   /organism="Drosophila melanogaster"
FT                   /strain="y2; cn bw sp"
FT                   /clone="BACR48K03 (D487) RPCI-98 48.K.3"
FT                   /clone_lib="RPCI-98 (Roswell Park Cancer Institute
FT                   Drosophila melanogaster BAC library, partial EcoRI in
```

```
FT                   pBACe3.6"
FT                   /map="57D3-57D12"
XX
SQ   Sequence 132637 BP; 35643 A; 30071 C; 30187 G; 36736 T; 0 other;


alignment_scores:
            Quality:  306.50          Length:    184
              Ratio:  2.620             Gaps:      7
 Percent Similarity:  63.587   Percent Identity: 38.043

alignment_block:
 US-09-060-609-2 x AC006092   ..

 Align seg 1/1  to: AC006092  from: 1  to: 132637

     107 SerLeuLysCysGluAspLeuLys...ValGlyGlnTyrIleCysLysAs 122
         ::::::   |||:::::::|||:::   :::||||||:::::|||   ||
    5182 AATGTGGACTGCAACGAGCTCCAGATGATGGGCCAGTTCATGTGCCCCGA 5231

     122 pPro.........LysIleAsnAspAlaThrGlnGluProValAsnCysT 136
         ||||         :::|||:::      ||||||:::   :::::::||||
    5232 CCCGGCGAGGGGTCAAATCGATCCCAAGACACAGCAGCTGGCCGGATGCA 5281

     136 hrAsnTyr...ThrAlaHisValSerCysPheProAlaProAsnIleThr 151
         ||:::         |||:::|||   |||:::   |||   :::|||:::
    5282 CCAGGGAGGGTCGCGCCCGCGTCTGGTGCATCGCCGCCAACGAGATCAAC 5331

     152 CysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGl 168
         |||   :::   :::||||||   |||
    5332 TGCACGGAG...ACTGGCAATGCCACG..................... 5355

     168 yPhePheLysProIleSerCysArg..AsnVal................ 178
          |||   :::   :::   |||:::  :::|||
    5356 .TTCACCCGAGAAGTGCCCTGCAAGTGGACGTGAGTGCAGCTACAGTATC 5404

     179 ..........................................AsnG 180
                                                    ||||
    5405 TCATAAACATAATCTTAAGACCTATATTCTATTCCTATTCTATAGGAATG 5454

     180 lyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeu 196
         |||||         :::::::::   ||||||:::|||||||||   :::
    5455 GCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTC 5504

     197 GlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPh 213
         |||:::|||||||||||||||||||||||:::::::|||||||||||||||
    5505 GGCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTT 5554

     213 eCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuI 230
         |||||||:::|||         :::|||:::|||||||||:::::::||||
    5555 CTGCACCCTCGGCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGA 5604

     230 leSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIleAsp 246
         ||:::::|||:::|||||||||:::||||||||:::|||:::|||
    5605 TAGCCCTGCAGGTTGTGGGTCCGGCGGATGGCTCCGCCTATGTGATACCC 5654

     247 TyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrPheAr 263
         |||||||||:::  :::      :::         |||   |||:::||
    5655 TACTACGGAGCCGGCATCCACATCGTGCGCAGCGACAACACCACGTACCG 5704

     263 g 263
         |
    5705 G 5705

seq_name: gb_htg2:AC005718

seq_documentation_block:
LOCUS       AC005718   149592 bp    DNA             HTG       30-JUL-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone DS02336 (D440) map
            60C8-60D2 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 68
            unordered pieces.
ACCESSION   AC005718
VERSION     AC005718.10  GI:5656710
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 149592)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149592)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Jul 30, 1999 this sequence version replaced gi:5630036.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases. P1 library location:
            25-32.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      343: contig of 343 bp in length
            *      344      423: gap of unknown length
            *      424     1169: contig of 746 bp in length
            *     1170     1249: gap of unknown length
            *     1250     2094: contig of 845 bp in length
            *     2095     2174: gap of unknown length
            *     2175     3405: contig of 1231 bp in length
            *     3406     3485: gap of unknown length
            *     3486     4245: contig of 760 bp in length
            *     4246     4325: gap of unknown length
            *     4326     5054: contig of 729 bp in length
            *     5055     5134: gap of unknown length
            *     5135     6031: contig of 897 bp in length
            *     6032     6111: gap of unknown length
            *     6112     7163: contig of 1052 bp in length
            *     7164     7243: gap of unknown length
            *     7244     8013: contig of 770 bp in length
            *     8014     8093: gap of unknown length
            *     8094     8951: contig of 858 bp in length
            *     8952     9031: gap of unknown length
            *     9032     9815: contig of 784 bp in length
            *     9816     9895: gap of unknown length
            *     9896    10572: contig of 677 bp in length
            *    10573    10652: gap of unknown length
            *    10653    11442: contig of 790 bp in length
            *    11443    11522: gap of unknown length
            *    11523    12232: contig of 710 bp in length
            *    12233    12312: gap of unknown length
```

```
*    12313    13278: contig of 966 bp in length
*    13279    13358: gap of unknown length
*    13359    14011: contig of 653 bp in length
*    14012    14091: gap of unknown length
*    14092    14897: contig of 806 bp in length
*    14898    14977: gap of unknown length
*    14978    15740: contig of 763 bp in length
*    15741    15820: gap of unknown length
*    15821    16793: contig of 973 bp in length
*    16794    16873: gap of unknown length
*    16874    17825: contig of 952 bp in length
*    17826    17905: gap of unknown length
*    17906    18911: contig of 1006 bp in length
*    18912    18991: gap of unknown length
*    18992    19720: contig of 729 bp in length
*    19721    19800: gap of unknown length
*    19801    20785: contig of 985 bp in length
*    20786    20865: gap of unknown length
*    20866    22392: contig of 1527 bp in length
*    22393    22472: gap of unknown length
*    22473    24290: contig of 1818 bp in length
*    24291    24370: gap of unknown length
*    24371    25694: contig of 1324 bp in length
*    25695    25774: gap of unknown length
*    25775    26731: contig of 957 bp in length
*    26732    26811: gap of unknown length
*    26812    27793: contig of 982 bp in length
*    27794    27873: gap of unknown length
*    27874    29161: contig of 1288 bp in length
*    29162    29241: gap of unknown length
*    29242    30491: contig of 1250 bp in length
*    30492    30571: gap of unknown length
*    30572    32263: contig of 1692 bp in length
*    32264    32343: gap of unknown length
*    32344    34053: contig of 1710 bp in length
*    34054    34133: gap of unknown length
*    34134    35477: contig of 1344 bp in length
*    35478    35557: gap of unknown length
*    35558    36683: contig of 1126 bp in length
*    36684    36763: gap of unknown length
*    36764    39017: contig of 2254 bp in length
*    39018    39097: gap of unknown length
*    39098    41095: contig of 1998 bp in length
*    41096    41175: gap of unknown length
*    41176    42262: contig of 1087 bp in length
*    42263    42342: gap of unknown length
*    42343    43644: contig of 1302 bp in length
*    43645    43724: gap of unknown length
*    43725    45000: contig of 1276 bp in length
*    45001    45080: gap of unknown length
*    45081    46626: contig of 1546 bp in length
*    46627    46706: gap of unknown length
*    46707    50898: contig of 4192 bp in length
*    50899    50978: gap of unknown length
*    50979    54635: contig of 3657 bp in length
*    54636    54715: gap of unknown length
*    54716    59775: contig of 5060 bp in length
*    59776    59855: gap of unknown length
*    59856    70727: contig of 10872 bp in length
*    70728    70807: gap of unknown length
*    70808   132766: contig of 61959 bp in length
*   132767   132846: gap of unknown length
*   132847   133588: contig of 742 bp in length
*   133589   133668: gap of unknown length
*   133669   134331: contig of 663 bp in length
*   134332   134411: gap of unknown length
*   134412   135076: contig of 665 bp in length
*   135077   135156: gap of unknown length
*   135157   135778: contig of 622 bp in length
*   135779   135858: gap of unknown length
*   135859   136505: contig of 647 bp in length
*   136506   136585: gap of unknown length
*   136586   137259: contig of 674 bp in length
*   137260   137339: gap of unknown length
*   137340   137991: contig of 652 bp in length
*   137992   138071: gap of unknown length
*   138072   138762: contig of 691 bp in length
*   138763   138842: gap of unknown length
*   138843   139393: contig of 551 bp in length
*   139394   139473: gap of unknown length
*   139474   140102: contig of 629 bp in length
*   140103   140182: gap of unknown length
*   140183   140811: contig of 629 bp in length
*   140812   140891: gap of unknown length
*   140892   141500: contig of 609 bp in length
*   141501   141580: gap of unknown length
*   141581   142262: contig of 682 bp in length
*   142263   142342: gap of unknown length
*   142343   143046: contig of 704 bp in length
*   143047   143126: gap of unknown length
*   143127   143812: contig of 686 bp in length
*   143813   143892: gap of unknown length
*   143893   144614: contig of 722 bp in length
*   144615   144694: gap of unknown length
*   144695   145373: contig of 679 bp in length
*   145374   145453: gap of unknown length
*   145454   146116: contig of 663 bp in length
*   146117   146196: gap of unknown length
*   146197   146808: contig of 612 bp in length
*   146809   146888: gap of unknown length
*   146889   147514: contig of 626 bp in length
*   147515   147594: gap of unknown length
*   147595   148161: contig of 567 bp in length
*   148162   148241: gap of unknown length
*   148242   148927: contig of 686 bp in length
*   148928   149007: gap of unknown length
*   149008   149592: contig of 585 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149592
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /clone="DS02336 (D440)"
                     /chromosome="2"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
                     /map="60C8-60D2"
BASE COUNT    37500 a  33853 c  33980 g  38891 t   5368 others
ORIGIN


alignment_scores:
            Quality:  257.00              Length:     85
              Ratio:  3.725                 Gaps:      0
 Percent Similarity:  81.176   Percent Identity: 54.118

alignment_block:
 US-09-060-609-2 x AC005718/rev   ..

 Align seg 1/1  to reverse of: AC005718  from: 1  to: 149592

     179 AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTr 195
         |||||||||         :::::::::   ||||||:::|||||||||
   25592 AATGGCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCAT 25543

     195 pLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuL 212
          :::|||:::||||||||||||||||||||||||||:::::|||||||||
   25542 GTTCGGCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCA 25493

     212 ysPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIle 228
         ||||||||||||:::|||         :::|||:::|||||||||:::::
   25492 AGTTCTGCACCCTCGGCGGCATGTTCCTGGGCCAGCTGATTGACATCGTG 25443

     229 LeuIleSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIl 245
         ||||||:::::::|||:::||||||||||:::|||||||||:::|||:::||
   25442 CTGATAGCCCTGCAGGTTGTGGGTCCGGCGGATGGCTCCGCCTATGTGAT 25393
```

```
   245 eAspTyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrP 262
       |   |||||||||:::   :::      :::         |||   |||:
 25392 ACCCTACTACGGAGCCGGCATCCACATCGTGCGCAGCGACAACACCACGT 25343

   262 heArg 263
       ::|||
 25342 ACCGG 25338

seq_name: gb_htg1:DMBR30C13

seq_documentation_block:
LOCUS       DMBR30C13  117674 bp    DNA             HTG       19-OCT-1999
DEFINITION  Drosophila melanogaster chromosome X clone BACR30C13 map 8D strain
            y; cn bw sp, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION   AL122025
VERSION     AL122025.1  GI:6090476
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 117674)
  AUTHORS   Madueno,E., de Pablos,B. and Modolell,J.
  TITLE     Sequencing the X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    Centro de Biologia Molecular Severo Ochoa, Consejo Superior de,
            Investigaciones Cientificas and Universidad Autonoma de Madrid,
            28049 Madrid, Spain.
REFERENCE   2  (bases 1 to 117674)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1999) European Drosophila Genome Sequencing
            Consortium
COMMENT     This is a 'working draft' sequence.  It currently
            consists of 153 contigs. The true order of the pieces is not known
            and their order in this sequence record is arbitrary. Gaps between
            the contigs are represented as runs of N, but the exact sizes of
            the gaps are unknown. This record will be updated with the finished
            sequence. 1      513: contig of 513 in length
            514      613: gap of unknown length
            614     1380: contig of 767 in length
            1381     1480: gap of unknown length
            1481     1995: contig of 515 in length
            1996     2095: gap of unknown length
            2096     2862: contig of 767 in length
            2863     2962: gap of unknown length
            2963     3754: contig of 792 in length
            3755     3854: gap of unknown length
            3855     4657: contig of 803 in length
            4658     4757: gap of unknown length
            4758     5330: contig of 573 in length
            5331     5430: gap of unknown length
            5431     5932: contig of 502 in length
            5933     6032: gap of unknown length
            6033     6474: contig of 442 in length
            6475     6574: gap of unknown length
            6575     7188: contig of 614 in length
            7189     7288: gap of unknown length
            7289     7513: contig of 225 in length
            7514     7613: gap of unknown length
            7614     8268: contig of 655 in length
            8269     8368: gap of unknown length
            8369     9346: contig of 978 in length
            9347     9446: gap of unknown length
            9447    10066: contig of 620 in length
            10067    10166: gap of unknown length
            10167    10814: contig of 648 in length
            10815    10914: gap of unknown length
            10915    11203: contig of 289 in length
            11204    11303: gap of unknown length
            11304    11941: contig of 638 in length
            11942    12041: gap of unknown length
            12042    12665: contig of 624 in length
            12666    12765: gap of unknown length
            12766    13410: contig of 645 in length
            13411    13510: gap of unknown length
            13511    13996: contig of 486 in length
            13997    14096: gap of unknown length
            14097    14677: contig of 581 in length
            14678    14777: gap of unknown length
            14778    15318: contig of 541 in length
            15319    15418: gap of unknown length
            15419    16047: contig of 629 in length
            16048    16147: gap of unknown length
            16148    16788: contig of 641 in length
            16789    16888: gap of unknown length
            16889    17502: contig of 614 in length
            17503    17602: gap of unknown length
            17603    18281: contig of 679 in length
            18282    18381: gap of unknown length
            18382    18819: contig of 438 in length
            18820    18919: gap of unknown length
            18920    19754: contig of 835 in length
            19755    19854: gap of unknown length
            19855    20699: contig of 845 in length
            20700    20799: gap of unknown length
            20800    21183: contig of 384 in length
            21184    21283: gap of unknown length
            21284    22021: contig of 738 in length
            22022    22121: gap of unknown length
            22122    23187: contig of 1066 in length
            23188    23287: gap of unknown length
            23288    23945: contig of 658 in length
            23946    24045: gap of unknown length
            24046    24616: contig of 571 in length
            24617    24716: gap of unknown length
            24717    25676: contig of 960 in length
            25677    25776: gap of unknown length
            25777    26190: contig of 414 in length
            26191    26290: gap of unknown length
            26291    27292: contig of 1002 in length
            27293    27392: gap of unknown length
            27393    28107: contig of 715 in length
            28108    28207: gap of unknown length
            28208    28729: contig of 522 in length
            28730    28829: gap of unknown length
            28830    29976: contig of 1147 in length
            29977    30076: gap of unknown length
            30077    30928: contig of 852 in length
            30929    31028: gap of unknown length
            31029    31903: contig of 875 in length
            31904    32003: gap of unknown length
            32004    32418: contig of 415 in length
            32419    32518: gap of unknown length
            32519    33132: contig of 614 in length
            33133    33232: gap of unknown length
            33233    33823: contig of 591 in length
            33824    33923: gap of unknown length
            33924    34827: contig of 904 in length
            34828    34927: gap of unknown length
            34928    35934: contig of 1007 in length
            35935    36034: gap of unknown length
            36035    36962: contig of 928 in length
            36963    37062: gap of unknown length
            37063    37679: contig of 617 in length
            37680    37779: gap of unknown length
            37780    38484: contig of 705 in length
            38485    38584: gap of unknown length
            38585    39541: contig of 957 in length
            39542    39641: gap of unknown length
            39642    40517: contig of 876 in length
            40518    40617: gap of unknown length
            40618    41307: contig of 690 in length
            41308    41407: gap of unknown length
            41408    42011: contig of 604 in length
```

```
    42012    42111: gap of unknown length
    42112    42828: contig of 717 in length
    42829    42928: gap of unknown length
    42929    43593: contig of 665 in length
    43594    43693: gap of unknown length
    43694    44391: contig of 698 in length
    44392    44491: gap of unknown length
    44492    45147: contig of 656 in length
    45148    45247: gap of unknown length
    45248    46798: contig of 1551 in length
    46799    46898: gap of unknown length
    46899    47830: contig of 932 in length
    47831    47930: gap of unknown length
    47931    48883: contig of 953 in length
    48884    48983: gap of unknown length
    48984    50341: contig of 1358 in length
    50342    50441: gap of unknown length
    50442    51797: contig of 1356 in length
    51798    51897: gap of unknown length
    51898    52237: contig of 340 in length
    52238    52337: gap of unknown length
    52338    53710: contig of 1373 in length
    53711    53810: gap of unknown length
    53811    54430: contig of 620 in length
    54431    54530: gap of unknown length
    54531    55595: contig of 1065 in length
    55596    55695: gap of unknown length
    55696    56555: contig of 860 in length
    56556    56655: gap of unknown length
    56656    57909: contig of 1254 in length
    57910    58009: gap of unknown length
    58010    59436: contig of 1427 in length
    59437    59536: gap of unknown length
    59537    60724: contig of 1188 in length
    60725    60824: gap of unknown length
    60825    62193: contig of 1369 in length
    62194    62293: gap of unknown length
    62294    63069: contig of 776 in length
    63070    63169: gap of unknown length
    63170    64233: contig of 1064 in length
    64234    64333: gap of unknown length
    64334    64782: contig of 449 in length
    64783    64882: gap of unknown length
    64883    66244: contig of 1362 in length
    66245    66344: gap of unknown length
    66345    67507: contig of 1163 in length
    67508    67607: gap of unknown length
    67608    68877: contig of 1270 in length
    68878    68977: gap of unknown length
    68978    70692: contig of 1715 in length
    70693    70792: gap of unknown length
    70793    71673: contig of 881 in length
    71674    71773: gap of unknown length
    71774    72809: contig of 1036 in length
    72810    72909: gap of unknown length
    72910    74298: contig of 1389 in length
    74299    74398: gap of unknown length
    74399    76722: contig of 2324 in length
    76723    76822: gap of unknown length
    76823    79379: contig of 2557 in length
    79380    79479: gap of unknown length
    79480    79862: contig of 383 in length
    79863    79962: gap of unknown length
    79963    80475: contig of 513 in length
    80476    80575: gap of unknown length

alignment_scores:
           Quality:  183.50              Length:    188
             Ratio:  1.799                Gaps:        9
 Percent Similarity:  54.255   Percent Identity: 32.979

alignment_block:
 US-09-060-609-2 x DMBR30C13    ..
```

```
Align seq 1/1  to: DMBR30C13  from: 1  to: 117674

     61 SerLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAlaVa 77
        |||   |||::::::||||||   :::|||   |||:::   :::   ::
   4022 TCAACAATGTCCAGTGCCTGGGCGAACGGAGTTTCCAGCGACAAATGAAT 4071

     77 lThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGly. 93
        :                                      :::   |||:::
   4072 T.................................GCCGATACTGCTA 4085

     94 ProTrpGlyAlaValAlaThr.SerAlaGlyGlyGluGluSerLeuLysC 110
        |||   |||   |||||||||| |||:::|||   :::::::|||
   4086 CCAAACGGAAATGTGGCAACAGAGCTGTGGCCAGCGCTCCAGT....... 4128

    110 ysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsn 126
                                                       |||
   4129 ........................................TGCAAC 4134

    127 AspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHis..... 141
        :::||||||:::::      :::   ||||||   |||:::|||
   4135 TCGGCGACGGATAAGCTATTCCGT...ACCAACTGCACCGTGCACCATGA 4181

    142 .ValSerCysPheProAlaProAsnIleThrCysLysAspSerSerGlyA 158
         |||   |||:::                                  ||||
   4182 TGTGCTCTGCCTC..............................GGCA 4198

    158 snGluThrHisPheThrGlyAsnGluValGlyPhePheLysProIleSer 174
        ||::::::   ||||||   |||                      :::
   4199 ATCGATCC...TTTACCCGCAAC.....................CTGCGA 4224

    175 CysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeuSerLe 191
        |||:::   ::::::||||||   :::::::::|||:::   :::|||||
   4225 TGCAACTGGACGCAGGGCTATCGGTGGAGCACAGCCCTGCTGATTAGCCT 4274

    191 uPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaL 208
        |   ||||||   :::||||||||||||||||||||||||:::
   4275 CACGCTAGGGGGCTTTGGAGCCGATCGATTCTACTTGGGCCACTGGCAGG 4324

    208 euGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGlySerLeu 224
          |||:::   |||:::   :::   |||      |||:::   ::::::
   4325 AAGGGATTGGAAAGCTGTTCAGCTTTGGCGGCCTCGGCGTCTGGACCATC 4374

    225 IleAspPheIleLeuIleSerMetGlnIleValGlyProSerAspGlySe 241
        ||||||   :::||||||||||||:::   :::|||||||:::||||||||
   4375 ATCGATGTCCTGCTCATCTCGATGCATTATTTGGGGCCAGCGGATGGCTC 4424

    241 rSerTyrIle 244
        |   ||||||
   4425 ACTTTACATA 4434

seq_name: gb_in2:AF181623

seq_documentation_block:
LOCUS       AF181623     1440 bp    mRNA            INV       16-SEP-1999
DEFINITION  Drosophila melanogaster BcDNA.GH02340 (BcDNA.GH02340) mRNA,
            complete cds.
ACCESSION   AF181623
VERSION     AF181623.1  GI:5901801
KEYWORDS    FLI_CDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1440)
  AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
            Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
            Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
            Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
            Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
```

```
  TITLE     Full Length Drosophila melanogaster cDNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1440)
  AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
            Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
            Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
            Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
            Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Berkeley Drosophila Genome Project,
            University of California Berkeley, 539 Life Sciences Addition
            #3200, Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            University of California Berkeley
            Berkeley, CA 94720
            For further information about this sequence,including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .1440
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH02974"
     gene            1. .1440
                     /gene="BcDNA.GH02974"
                     /note="motif=(desc:'C_TYPE_LECTIN_1', acc:'PS00615',
                     count:'1', method:'PPSEARCH', dbase:'PROSITE')"
     CDS             276. .1130
                     /gene="BcDNA.GH02974"
                     /codon_start=1
                     /product="BcDNA.GH02974"
                     /protein_id="AAD55409.1"
                     /db_xref="GI:5901802"
                     /db_xref="FLYBASE:FBcd0002974"
                     /translation="MRLQRQCIVVNMRSAIVLIMIFVLTGIRNSETASGGNQMDLSDS
                     KGDHKDNSNASNGNGNANDNEVYVPPLVSSMVAKSGGGAGGLLDNITAYSSSSSSSSS
                     NGNNNMLCPYDKETPCDRLQFPCIRCNYNHGCIYGRDLNVTCEVINNVQCLGERSFQR
                     QMNCRYCYQTEMWQQSCGQRSSCNSATDKLFRTNCTVHHDVLCLGNRSFTRNLRCNWT
                     QGYRWSTALLISLTLGGFGADRFYLGHWQEGIGKLFSFGGLGVWTIIDVLLISMHYLG
                     PADGSLYI"
BASE COUNT      422 a    325 c    329 g    364 t
ORIGIN

alignment_scores:
           Quality:  183.50            Length:    188
             Ratio:  1.799               Gaps:      9
 Percent Similarity:  54.255   Percent Identity: 32.979

alignment_block:
 US-09-060-609-2 x AF181623   ..

 Align seg 1/1  to: AF181623  from: 1  to: 1440

      61 SerLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAlaVa 77
         |||   |||::::::|||||   :::|||   |||:::   :::   ::
     715 TCAACAATGTCCAGTGCCTGGGCGAACGGAGTTTCCAGCGACAAATGAAT 764

      77 lThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGly. 93
         :                                 :::   |||:::
     765 T.................................GCCGATACTGCTA 778

      94 ProTrpGlyAlaValAlaThr.SerAlaGlyGlyGluGluSerLeuLysC 110
         |||   |||   |||||||| |||:::|||   ::::::|||
     779 CCAAACGGAAATGTGGCAACAGAGCTGTGGCCAGCGCTCCAGT....... 821

     110 ysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsn 126
                                                     |||
     822 ........................................TGCAAC 827

     127 AspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHis..... 141
         :::||||||::::::      :::   ||||||   |||:::|||
     828 TCGGCGACGGATAAGCTATTCCGT...ACCAACTGCACCGTGCACCATGA 874

     142 .ValSerCysProAlaProAsnIleThrCysLysAspSerSerGlyA 158
          |||   |||:::                                 ||||
     875 TGTGCTCTGCCTC.............................GGCA 891

     158 snGluThrHisPheThrGlyAsnGluValGlyPhePheLysProIleSer 174
         ||::::::   ||||||   |||                     :::
     892 ATCGATCC...TTTACCCGCAAC.....................CTGCGA 917

     175 CysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeuSerLe 191
         |||:::   ::::::||||||   :::::::::|||:::   :::||||||
     918 TGCAACTGGACGCAGGGCTATCGGTGGAGCACAGCCCTGCTGATTAGCCT 967

     191 uPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaL 208
         |   ||||||   :::||||||||||||||||||||||||||:::
     968 CACGCTAGGGGGCTTTGGAGCCGATCGATTCTACTTGGGCCACTGGCAGG 1017

     208 euGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGlySerLeu 224
           |||:::   |||:::   :::   |||      |||:::   ::::::
    1018 AAGGGATTGGAAAGCTGTTCAGCTTTGGCGGCCTCGGCGTCTGGACCATC 1067

     225 IleAspPheIleLeuIleSerMetGlnIleValGlyProSerAspGlySe 241
         ||||||   :::||||||||||||:::   :::||||||:::||||||||
    1068 ATCGATGTCCTGCTCATCTCGATGCATTATTTGGGGCCAGCGGATGGCTC 1117

     241 rSerTyrIle 244
         |   ||||||
    1118 ACTTTACATA 1127

seq_name: gb_in2:CELC41D11

seq_documentation_block:
LOCUS       CELC41D11   40912 bp    DNA             INV       27-FEB-1998
DEFINITION  Caenorhabditis elegans cosmid C41D11.
ACCESSION   AF003740
VERSION     AF003740.1  GI:2105490
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40912)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 40912)
  AUTHORS   Gattung,S and Maggi,L.
  TITLE     The sequence of C. elegans cosmid C41D11
  JOURNAL   Unpublished (1998)
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO  63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5′ cosmid is C11D9, 3100 bp overlap. Actual start of this cosmid is at base position 1 of CELC41D11; actual end is at 40912 of CELC41D11

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers

source 1. .40912
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="C41D11"

gene complement(6009. .8933)
/gene="C41D11.3"

CDS complement(join(6009. .6173,6597. .6821,7005. .7699, 7830. .8157,8748. .8933))
/gene="C41D11.3"
/note="weak similarity to rat mitochondrial capsule selenoprotein (GB:X87883) and integrin beta oligodendroglia chain precursor (NID:g520566); coded for by C. elegans cDNA yk2d8.3; coded for by C. elegans cDNA yk130f5.3; coded for by C. elegans cDNA yk101b2.3; coded for by C. elegans cDNA yk121e1.3; coded for by C. elegans cDNA yk130f5.5; coded for by C. elegans cDNA yk101b2.5; coded for by C. elegans cDNA yk121e1.5"
/codon_start=1
/protein_id="AAC48138.1"
/db_xref="GI:2105493"
/translation="MVPLVKKQMSIDEKIMEDDEDEGFGDSMKTIKRGQGSSLARSLS LKLSPKTDQQKPIGADTEKSCSTPSLPTAPLLSRQGSTSSLRFKVSPSTFLKPIATPI IDVSPTPRPDPEPLRIIVAAAAAAPVESTELSEPSTSKTPEQKAEPAMQKSALRKHAA NFSVREEIFIPGIRRKKIRFSGLHVHYFDRRQGGSTVPTEGDISLDMHNKHHSHRYFS LSTGKRPILDLLLYADDDLSDGEVIDPDSDEDRDYDEYASAKSIPRINQRHRIKLLKK SGVKVEKNTDAIDSLRNTRVECGCSCENGVCLPETCQCAIDGIHCQVDGGEWPTQPCA CFAETCTNPEGRTFYDPEAVHEFRQRTIMNWRASQKTGISGSPVVKKFADSDEEDDKA KTWLLKNAIPIKLEESPTKYPVTPVYTRRSRSSLSDIRESTSETTGSITPSTSLSERI KAMQHLDNEVCVEQEEKDEDDAYFMDNKEDTIIEADLTLQEVYVEQDMVAVTVESSSE TTDIDANEMIEATAMSPVKTTLVV"

gene complement(18663. .20974)
/gene="C41D11.4"

CDS complement(join(18663. .18779,18829. .18944,19622. .19868, 20561. .20974))
/gene="C41D11.4"
/note="coded for by C. elegans cDNA yk21e1.5; coded for by C. elegans cDNA yk2d8.5"
/codon_start=1
/protein_id="AAC48139.1"
/db_xref="GI:2105494"
/translation="MVSVAASLKRAGVPSLTRTKTSALPIKMEKKPMVPLRIEQDHQT PIRKLAPESIPSLEDQVTPPMRRSSSGSMLLRQTPERIRRQEEDVETPGSPVKRYRRS GSMPRLEAQSLDVPIRMEIGNSLTPVLKRKRTSSLTCLFSSDMPPILTPQRLRPRGIS ETRTIDDDVFVPKSPTKKPITSPIKKRVRQDSENRSTSPLKKRYISSFSIRLEEEKEI EKAVPTPTMSTPMRKFKSTIFNPRAVTERIAQTYGTGSRQLLKSTSSTVLPAPPALVA RSTSGETYYSATEGGLKFFKY"

gene 21167. .23273
/gene="C41D11.2"

CDS join(21167. .21266,21736. .21863,21909. .22031,22078. .22273, 22321. .22420,22470. .22941,23007. .23273)
/gene="C41D11.2"
/note="coded for by C. elegans cDNA yk83c1.3; coded for by C. elegans cDNA yk83c1.5"
/codon_start=1
/protein_id="AAC48137.1"
/db_xref="GI:2105492"
/translation="MVSETNLQLFFNMHLVIILNLQMDSNKASLFKNWLSKMSTAVTI TAPSVKHILLDSLVVMKIVKHVDSELHAGISEVSGDACAGVLTGLVFLEDSRLEITNC FPTVRNEPVMDDDANAAQQYEEQKQHEMLDMLRKFRTMNIDYEIVGFYQSHQFGAGFS HDLVESMFDYQAMGPENVVLIYDPIKTRQGQLSLRAWRLSTAALDLASKNDWRPELVK AAGLTYQNMFEELPIIIKSSYLNNVLMSELSLAKSCSSDKYSTRHFDLGSKKSLEKSV RAMMANVDELNKSIQSLTKYTIDKQRHDNMVFSLTQKRVRISNFEIKNILIVSLKFIF RSVTAFKPKVLFFENHFPTYRGSTNYKTGCEGYIFGGIKQQENESRVARGDPTIPMDD IKRIKAPQLQTRNGLLDELLASFDTNALADFSKTVTSENITKMFIAEAVAEEKVAGTK DRTLSSVSSTR"

gene complement(23679. .27940)
/gene="C41D11.5"

CDS complement(join(23679. .24030,24286. .24440,24497. .24619, 25129. .25305,25352. .25478,25755. .26011,26081. .26330, 26721. .26791,27033. .27433,27495. .27600,27647. .27731, 27783. .27940))
/gene="C41D11.5"
/note="similar to the DNA/RNA non-specific endoneuclease family; coded for by C. elegans cDNA CEESU20R"
/codon_start=1
/protein_id="AAC48141.1"
/db_xref="GI:2105496"
/translation="MMSLSSLNNENLNDEIIDFESKCKRFPVMFKRAVCIHSSLQFVD LFMQNLMFSHGFSSTPIIARQEDETEVLSKTMLSFYKIRNEFREIFRSHHRGNVSEVA IFIGTCPVRSTYSYRVPITICEAEDSTHKSCGDTCAELNDFERRKINRDLFMHSCSKE QQSNETNKTNKNHRLFVFIKGSEDMVNEEIEEELVHEFPPNFFFLFFFFRDPPFGSKQ SFHYSFIHEKCNCDGVPAPRNDGKWLRMVPYIMLHKILFLICLASFIPTIGSISGTKD VKSKNCDGSAGLTCTFPGDCRIGDTVKVNCTSRKGCPNPVSRNNVEAVCRFCWQLLPG DYDCEPATNCSTSSTKLLVTKCSAHSSVICMGQRNFYKRIPCNWSSGYSWTKTMILSV VLGGFGADRFYLGLWKSAIGKLFSFGGLGVWTLVDVVLIAVGYIKPYDGSIRNMIGKV AGTAAIAGISFLAGKYSNDDLPIFRNVQSATNVPMNQIQVSEPMTVKPASLNADAMGP SRSAEIMKHGYPGFTNVRTYEDFVLSYDYKTRTAHWVCEHLTPERLKHAEGVDRKLCE FKPDITFPQKFLSQNTDYKCSGFDRGHLAAAGNHRKSQLAVDQTFYLSNMSPQVGRGF NRDKWNDLEMHCRRVAKKMINSYIITGPLYLPKLEGDGKKYIKYQVIGDNNVAVPTHF FKVALFEVTPGKFELESYILPNAVIEDTVEISKFHVPLDAVERSAGLEIFARLDPKSI VKENGAKKGGLLW"

gene complement(28522. .33822)
/gene="C41D11.7"

CDS complement(join(28522. .28843,29286. .29418,29463. .29705, 30317. .30489,30536. .30718,31647. .31861,31908. .32009, 32388. .32461,32961. .33352,33758. .33822))
/gene="C41D11.7"
/note="similar to mammalian DNA-binding proteins SMUBP-2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC48142.1"
/db_xref="GI:2105497"
/translation="MIEADKKFLELVKKWGVLSVGLNQKQNTNPVYQYTLEDNPESLT LQDVLESHMTDPIELNSFDSMEKEAGIPFSNEKQRTAIKMALNENRKLVCIQGPPGTG KTFTLTLLLCRLIQQKKQVVVLAPTREALANIRMMTKKTLKRMGIKVHEHALMDTNEY RDVINKSDRALMAAEEVRDLRKAFDNGEITENVLDEMRQSIINRVRNEVGAEVIGNVR VAFATIGASFVDFVMKHKKFDPCLCIIDEAAQVMEAQTWPAVYKMKRIVMAGDPKQLP ALVFTDEAKAFGLQNSVMDRILEKKNNFSWIMLENQYRSNAKIATWSNTCFYHNQLKT DVKCHEYSLHTILNPQPKKFRNLFDPLVLIDTSLERDVEKRLETYEHAVFDTNSINKT KQGFSYANLAEAKIAIGHYQRLLKYGVQPSDIAIITPYKGQTSLVTKLMEEFGAETGY TDFVQTTIGTVDSVQGKEYEVVIFTMVRSNPRKTMGFVSELRRLNVVITRAKRHFMFI GNGYLLAESKKDEIRKLYDCFKNAKRRFHPNNAFGEDGNVTDYVSNNFGQDLESFMNY SNDEEMINWCKRFNENGPDYRAKKALAWEKREETRLMKTISKVKYEYKRGSTEQLASE LDQLDCSSN"

gene 34408. .35688
/gene="C41D11.1"

CDS join(34408. .34592,34661. .34754,34805. .34917,34967. .35153, 35210. .35348,35399. .35505,35551. .35688)
/gene="C41D11.1"

```
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC48136.1"
                     /db_xref="GI:2105491"
                     /translation="MTDKENFYSGFFLGRLPGSFDSRFLIKVACFQPGKEERNVFELT
                     TDTKREEDYGEMVHFWVDKGPLQNKSFVRFEENYETFFIHENKMFDRIVHTDKDFYVG
                     VNDLPDSDIYIIHPDLGKMIDTRSIVTPGKKYNFITFEVKQNENNSFDIEIVGGEVQGK
                     EKNSEEPSAKEKIIQKMEQVHEVMKVEYEYNKNEETANWDVIGHLKLIKFYDKDLTSG
                     LRPVLRLMDNRHVALFKKQRETMPIKMIKVIEVDGIWKDDVNFVPQEGTLLPTDHDEK
                     TVVLLGNDYRHKFERLFKREDVFRLESNNTLTHLKFWQYGLEKI"
     gene            complement(37163. .39721)
                     /gene="C41D11.6"
     CDS             complement(37163. .39721)
                     /gene="C41D11.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC48140.1"

gnment_scores:
          Quality:  162.00            Length:    159
            Ratio:  2.051               Gaps:      5
  Percent Similarity:  49.686   Percent Identity: 28.931

alignment_block:
  US-09-060-609-2 x CELC41D11/rev   ..

  Align seg 1/1  to reverse of: CELC41D11  from: 1  to: 40912

      110 CysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAs 126
          |||   :::|||      |||:::|||   |||   :::|||      ||
    26170 TGTTGGCAACTTCTACCTGGGGATTATGATTGC...GAGCCAGCGACAAA 26124

      126 nAspAlaThrGlnGluPro.............ValAsnCysThr..... 136
          |   :::|||::::::                 ::::::|||:::
    26123 TTGCTCCACGAGTAGTACAAAGCTTCTCGTGACAAAGTGCTCGGTGAGCA 26074

      136 ................................................... 136

    26073 ATTTCTTTAATTTAATATCTGTTTTAAAATACCTAATTTTCAAAAAAAAA 26024

      137 ...AsnTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCy 152
             |||:::   ||||||   |||                  :::   ||
    26023 TTGAATTTTCAGGCTCACAGCTCA.................GTAATATG 25992

      152 sLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyP 169
          |                                 ::::::   :::|
    25991 TATG.............................GGACAACGTAATT 25975

      169 hePheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysVal 185
          ||:::|||   |||   |||:::      :::||||||||||:::
    25974 TCTACAAGCGAATTCCATGTAACTGGTCATCTGGATACAGTTGGACAAAA 25925

      186 AlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTy 202
          ::::::   |||||||:::   ||||||   :::||||||||||||||||
    25924 ACAATGATTCTCTCAGTTGTGCTCGGCGGTTTTGGAGCTGATCGTTTCTA 25875

      202 rLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheC 219
          |||||||            ::::::   |||:::   :::   |||
    25874 CTTGGGACTCTGGAAATCTGCAATTGGAAAACTTTTCAGTTTCGGTGGTC 25825

      219 ysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleVal 235
            |||:::   :::|||:::|||   :::||||||::::::      :::
    25824 TTGGTGTGTGGACTCTTGTGGACGTTGTTCTGATCGCAGTTGGATACATT 25775

      236 GlyProSerAspGlySerSerTyrIle 244
          |||   |||||||||   ||||||
    25774 AAACCTTACGATGGATCCATGTATATT 25748

seq_name: gb_htg6:AC016672

seq_documentation_block:
LOCUS       AC016672   205157 bp    DNA             HTG       04-DEC-1999
DEFINITION  Homo sapiens clone RP11-14C10, WORKING DRAFT SEQUENCE, 1 unordered
            pieces.
ACCESSION   AC016672
VERSION     AC016672.1  GI:6524410
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 205157)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205157)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     Center project name: H_NH0014C10.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   205157: contig of 205157 bp in length.
FEATURES            Location/Qualifiers
     source         1. .205157
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="RP11-14C10"
BASE COUNT    53433 a  46205 c  47212 g  58163 t    144 others
ORIGIN


alignment_scores:
          Quality:  139.00            Length:    130
            Ratio:  2.138               Gaps:      4
  Percent Similarity:  50.000   Percent Identity: 32.308

alignment_block:
 US-09-060-609-2 x AC016672   ..

 Align seg 1/1  to: AC016672  from: 1  to: 205157

      138 TyrThrAlaHisValSerCysPheProAlaProAsn.............. 149
          |||     |||:::::::|||:::|||||||   |||
    15567 TACTTTTGACATTTAAACTGTCTTCCTGCCCGTAATTATTTGGTGGTGTC 15616

      150 .............IleThrCysLysAspSerSerGlyAsnGluThrHisP 162
                      :::||||||         :::|||   :::
    15617 AGTTTGTGAAATAGTCACTTGTGTTCCAGTTAAGGGCTACGATCAGGGG. 15665

      162 heThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVal 178
                                  |||||||:::       ::::::
    15666 .........................AAGCCCTTTCTCAAATCTGACATA 15689

      179 Asn.............................GlyTyrSerTyrLysVa 185
                                         |||:::   ::::::
    15690 ATGTGGTGATTAAACCGTGACTGGCCACAGCTGGGGTGGAGGTGGTCTCA 15739

      185 lAlaValAla.............LeuSerLeuPheLeuGlyTrpLeuGlyA 198
           ::::::                 |||:::   ||||||   :::||||
    15740 TGGAATCCTGTCTGTTTGCCATCGCAGCATCACCCTCGGTGGGTTTGGAG 15789

      198 laAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCys 214
          ||||||||||||||||||            ||||||   |||:::
    15790 CAGACCGTTTCTACCTGGGCCAGTGGCGGGAAGGCCTCGGCAAGCTCTTC 15839

      215 ThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSe 231
          :::   |||      ||||||   :::|||||||||   :::|||||::
```

```
   15840 AGCTTCGGTGGCCTGGGAATATGGACGCTGATAGACGTCCTGCTCATTGG 15889

     231 rMetGlnIleValGlyProSerAspGlySerSerTyrIle 244
         ::::      ||||||||||:::||||||||||   ||||||
   15890 AGTTGGCTATGTTGGACCAGCAGATGGCTCTTTGTACATT 15929

seq_name: gb_in1:CELC02F5

seq_documentation_block:
LOCUS       CELC02F5    22333 bp    DNA             INV       22-OCT-1993
DEFINITION  C. elegans cosmid C02F5.
ACCESSION   L14745 L18807
VERSION     L14745.1  GI:289607
KEYWORDS    GRR1 gene; GTP-binding protein; glucose repression; proteasome
            component C5.
SOURCE      Caenorhabditis elegans (strain Bristol N2) hermaphrodite mixed
            whole animal DNA.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 22333)
  AUTHORS   Sulston,J., Du,Z., Thomas,K., Wilson,R., Hillier,L., Staden,R.,
            Halloran,N., Green,P., Thierry-Mieg,J., Qiu,L., Dear,S.,
            Coulson,A., Craxton,M., Durbin,R., Berks,M., Metzstein,M.,
            Hawkins,T., Ainscough,R. and Waterston,R.
  TITLE     The C.elegans genome sequencing project: A beginning
  JOURNAL   Nature 356, 37-41 (1992)
  MEDLINE   92168156
REFERENCE   2  (bases 1 to 22333)
  AUTHORS   Anderson,K.
  TITLE     Sequence of the C. elegans cosmid C02F5
  JOURNAL   Unpublished (1993)
COMMENT     Submitted by:
            Nematode Sequencing Project
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
            MRC Laboratory of Molecular Biology,
            Cambridge CB2 2QH, England
            e-mail: rw@nematode.wustl.edu and jes@cele.mrc-lmba.cam.ac.uk NOTE:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder (P. Green and L. Hillier, in preparation)
            Neighboring cosmid information:
            Bases 1 to 800 of this entry correspond to bases 26944 to 27743 of
            GenBank entry CELC30A5 (L10990).  Bases 22034 to 22333 of this
            entry
            correspond to bases 1 to 300 of GenBank CELF09G8 (L11247).  The
            sequence
            of C. elegans cosmid C02F5 begins at position 18944 of GenBank
            CELC30A5
            (L10990), continues through this entry and ends at position 15818
            of
            GenBank CELF09G8 (L11247).  The sequence of C. elegans cosmid F09G8
            begins at position 22034 of this entry.
FEATURES              Location/Qualifiers
     source           1. .22333
                      /organism="Caenorhabditis elegans"
                      /strain="Bristol N2"
                      /db_xref="taxon:6239"
                      /dev_stage="mixed"
                      /sex="hermaphrodite"
                      /tissue_type="whole animal"
                      /map="3"
     CDS              join(8. .124,175. .257,305. .694,742. .943,993. .1493,
                      2247. .2423)
                      /standard_name="C02F5.7 protein"
                      /note="homology with glucose induced repressor, GRR1;
                      putative"
                      /codon_start=1
                      /protein_id="AAA27922.1"
                      /db_xref="GI:289614"
                      /translation="METAERISAAASAASSRRAKRLAQQAHKTHPVIQAKQNQMYLIT
                      TLSPAQVDNSLINRVLPKEVLLKVFSFLDTKALCRSAQVCRSWSILALDGSNWQRVDL
                      FTFQRDVKTAVVENLARRCGGFLKELSLKGCENVHDSALRTFTSRCPNLEHLSLYRCK
                      RVTDASCENLGRYCHKLNYLNLENCSSITDRAMKYIGDGCPNLSYLNISWCDAIQDRG
                      VQIILSNCKSLDTLILRGCEGLTENVFGSVEAHMGAIKKLNLLQCFQLTDITVQNIAN
                      GATALEYLCMSNCNQISDRSLVSLGQHSHNLKVLELSGCTLLGDNGFIPLARGCRQLE
                      RLDMEDCSLISDHTINSLANNCTALRELSLSHCELITDESIQNLASKHRETLNVLELD
                      NCPQLTDSTLSHLRHCKALKRIDLYDCQNVSKEAIVRFQVYVDAVLSYDCIYVPCSTR
                      STENWYKLTKLTKSDINRAPCAHRVEAQLWYSIKNVLKF"
     exon             175. .257
                      /note="putative"
                      /number=9
     exon             305. .694
                      /note="putative"
                      /number=10
     exon             742. .943
                      /note="putative"
                      /number=11
     exon             993. .1493
                      /note="putative"
                      /number=12
     misc_feature     2786
                      /note="possible trans-spliced leader site for C02F5.6"
     CDS              join(2793. .2834,3022. .3473,3528. .3686,3737. .4091,
                      4139. .4252,4326. .4463)
                      /standard_name="C02F5.6 protein"
                      /note="putative"
                      /codon_start=1
                      /protein_id="AAA27921.1"
                      /db_xref="GI:289613"
                      /translation="MAHTSDGWGAPYDNQTYVEAYEQLEIALLEPLDRILETANVEEF
                      RPKFKNHQDPNDRKNKKNNDEEWRDSIYNIATDESDTDDHEQRKNFFQPPLQVQRNSF
                      VKNTLMEFKRSSQIDISRLAVMGCGEMSLEKGICEYLGSFGTINVLSVDIDEPSLSIG
                      QQLLGKAESIFLRKHLERNAEILAVETGLPVLMRSYVGDILEPDHRFADVDAIVSMEV
                      VEHIPLPNAKKFVENVLGTLMPRIFIFSTPNHEYNAVFGMEPGEFRHGDHKFEMNRKE
                      FSNWLEELSIRFPHYQIDPPHYIGMTRGYENLSGASQAAVCRLQVDLNTTLPQEVTPY
                      EMVGHLPCRLGSRLIAYNLVKEAFLDWLEKIELQEHEPRTDGYSPYWIFNVQNILHHL
                      KAPVSFALTIDEKVAMQVSSVIGTVEF"
     exon             3022. .3473
                      /note="putative"
                      /number=13
     exon             3528. .3686
                      /note="putative"
                      /number=14
     exon             3737. .4091
                      /note="putative"
                      /number=15
     exon             4139. .4252
                      /note="putative"
                      /number=16
     CDS              join(6347. .6436,6481. .6795,6876. .6992)
                      /standard_name="C02F5.5 protein"
                      /note="putative"
                      /codon_start=1
                      /protein_id="AAA27920.1"
                      /db_xref="GI:289612"
                      /translation="MSYMEQIPPYEYYKYVFGTRCAVLACSSFELMMVLLGSLGDSNV
                      LAKLFYLVFLGASAAVSGFNLQQNIDGRDEIKKITGSGDNETRGKAAALVGIPALAAV
                      FVFLCVSGHAFFSLFVLVHVLASIGQLGIEAYEVRNGFCIKAPENKKAVADSYDQNYQ
                      ILNMVQDSVFDKK"
     exon             6481. .6795
                      /note="putative"
                      /number=17
     CDS              complement(join(8274. .8477,8518. .8583,8631. .8897,
                      9041. .9178,9473. .9532))
                      /standard_name="C02F5.8 protein"
                      /note="putative"
                      /codon_start=1
                      /protein_id="AAA27923.1"
                      /db_xref="GI:289615"
                      /translation="MATWKFIIRSVLFFLDLAMLLAALALIAVGFWMGYDSSFDTDLK
                      NVIYKYDDPKSLADAKFNIRVWLIVVFWSIIGLSLGAVVTAVLGMISSVWPKRKGFMI
                      TYLVLIIVLVSLEIGCGVAVLVRRNSLHDNTNSLIDAMYTTNSVNDLKIIQDKYNCCG
                      IENSLFNVMYCGPMSQKPHCDVAVFDSVDNTMMISGIILLVILILQTIAIILPVPILI
                      SRKKTYKYSYEPRVTQLADITEDTRF"
     exon             8518. .8583
```

```
                /note="complement; putative"
                /number=6
     exon       8631. .8897
                /note="complement; putative"
                /number=7
     exon       9041. .9178
                /note="complement; putative"
                /number=8
     misc_feature 9558
                /note="possible trans-spliced leader site for C02F5.8"
     CDS        complement(join(11079. .11495,11809. .12165,12215. .12369,
                13369. .13474,13523. .13916,13962. .14227))
                /standard_name="C02F5.9 protein"
                /note="homology with proteasome component C5; coded for by
                C. elegans cDNAs GenBank:M80127 and T01498; putative"
                /codon_start=1
                /protein_id="AAA27924.1"
                /db_xref="GI:289616"
                /translation="MTSFTGITAVANATNEMAMFKQAMKEVAAHPEWMSSRQIERQRW
                NPYSMEGGSTCAISGENFAIVASDTRMTQNDINILTRDAEKIQILNDNIILTTSGFYG
                DVLQLKKVLQSRLHKYRFDYRSDMSVDLCAELLSRNLYYRRFFPYYTGAILAGIDEHG
                KGAVFSYDPIGCIERLGYSASGAAEPMIIPFLDCQIGHVTLSEGYERPELTLDRAISL
                MKDSFRGAAERE ISTGDKIHLVIAEAGKPVVVKFLPLQSISAQSQNVLKKPFKSISAE
                LGFAQPLNLVDTLGLKPRDSDAMSRGKRPASVLKGLFDDWDPTANLDIPGTSSDIPSD
                PSSALKVPKKEVLDESEEILDQTSGSSSFSLNDSEQANEQERNIIEDLLRQQMFTESK
                IKNEEASLLKGLKKGLGEELESSHANYVITCNFPGCGLKYNWRVKYGKLRLLDHALTH
                SNRKIPCKLCGFECTNVRRMRSHYAKAHPNERVEGYGMKALVSGDSSRIGDGVDGDID
                QQVSDDELKELWNNCFSESIHLVGQASGFVEGEKYRRMTKRRKLDREAMNSMNFMF"
     exon       11809. .12165
                /note="complement; putative"
                /number=2
     exon       12215. .12369
                /note="complement; putative"
                /number=3
     exon       13369. .13474
                /note="complement; putative"
                /number=4
     exon       13523. .13916
                /note="complement; putative"
                /number=5
     CDS        join(14580. .14717,14761. .14894,14937. .15120,15330. .15412,
                15465. .15612,15661. .15759)
                /standard_name="C02F5.4 protein"
                /note="coded for by C. elegans cDNA GenBank:CEL16D4
                (M89109); putative"
                /codon_start=1
                /protein_id="AAA27919.1"
                /db_xref="GI:289611"
                /translation="MADFTEQTLRQKLTNLSNHPSSIQTTSAWLLQNHSNRELIIRVW
                LKTVKKETHGSKIVNLLYVANDVSQNARKTCPQFKDEFFPAIESSFRHAIELKNAKEV
                EHAIGKLINVWKDRQIFTPSQCKRLHEVHQQVKLSGSFPTPAVANKEHGKSVRSSQEA
                QSLLVQNEEALPMLEEYVKRLKNETNERETLESNLNMLIENVRMSIEHHEKLCREVKR
                REDRIKADLLEVEKTFESLPDLAAEMPNAPLPTLEALFEKRKK"
     exon       14761. .14894
                /note="putative"
                /number=18
     exon       14937. .15120
                /note="putative"
                /number=19
     exon       15330. .15412
                /note="putative"
                /number=20
     exon       15465. .15612

alignment_scores:
           Quality:  138.50            Length:     78
             Ratio:  2.473               Gaps:      1
 Percent Similarity:  71.795   Percent Identity: 38.462

alignment_block:
 US-09-060-609-2 x CELC02F5   ..

 Align seg 1/1  to: CELC02F5  from: 1  to: 22333

     179 AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTr 195
         ||||||:::   :::   :::::::::      |||:::|||||||||
   18579 AATGGTCACTACTTTCTCACAACTCTTCTGTATTCCATTTTTCTCGGTGT 18628

     195 pLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuL 212
         :::::::::|||||||||   |||||||||   |||::::::::::   |
   18629 GGTTGCTGTCGACCGTTTTTGCCTGGGATATTCTGCAATGGCTGTCGGAA 18678

     212 ysPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIle 228
         ||:::   |||:::|||      ||||||      ::::::|||::::::
   18679 AACTAATGACTTTGGGTGGATTTGGAATCTGGTGGATTGTCGATATATTC 18728

     229 LeuIleSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIl 245
         |||:::   :::   ::::::|||||||:::|||   |||||||:::
   18729 TTACTGGTACTTGGAGTACTCGGACCCGCGGATGACTCGAGTTGG...GA 18775

     245 eAspTyrTyrGlyThrArgLeuThrArgLeuSer 256
         ||||||      :::|||||||:::   |||
   18776 GCCGTACTATTAAAAGAAATTAACTAATACGAGT 18809

seq_name: gb_htg6:AC011069

seq_documentation_block:
LOCUS       AC011069   168266 bp    DNA             HTG       02-DEC-1999
DEFINITION  Drosophila melanogaster chromosome X clone BACR11H20 (D881) RPCI-98
            11.H.20 map 12B-12C strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 92 unordered pieces.
ACCESSION   AC011069
VERSION     AC011069.6  GI:6503260
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 168266)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168266)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Dec 2, 1999 this sequence version replaced gi:6492479.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 92 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
* be preserved.
*       1      355: contig of 355 bp in length
*     356      435: gap of unknown length
*     436      758: contig of 323 bp in length
*     759      838: gap of unknown length
*     839     1738: contig of 900 bp in length
*    1739     1818: gap of unknown length
*    1819     2799: contig of 981 bp in length
*    2800     2879: gap of unknown length
*    2880     3599: contig of 720 bp in length
*    3600     3679: gap of unknown length
*    3680     4328: contig of 649 bp in length
*    4329     4408: gap of unknown length
*    4409     4905: contig of 497 bp in length
*    4906     4985: gap of unknown length
*    4986     5626: contig of 641 bp in length
*    5627     5706: gap of unknown length
*    5707     6238: contig of 532 bp in length
*    6239     6318: gap of unknown length
*    6319     6946: contig of 628 bp in length
*    6947     7026: gap of unknown length
*    7027     8149: contig of 1123 bp in length
*    8150     8229: gap of unknown length
*    8230     8867: contig of 638 bp in length
*    8868     8947: gap of unknown length
*    8948     9736: contig of 789 bp in length
*    9737     9816: gap of unknown length
*    9817    10567: contig of 751 bp in length
*   10568    10647: gap of unknown length
*   10648    11547: contig of 900 bp in length
*   11548    11627: gap of unknown length
*   11628    12564: contig of 937 bp in length
*   12565    12644: gap of unknown length
*   12645    13378: contig of 734 bp in length
*   13379    13458: gap of unknown length
*   13459    14616: contig of 1158 bp in length
*   14617    14696: gap of unknown length
*   14697    15356: contig of 660 bp in length
*   15357    15436: gap of unknown length
*   15437    16193: contig of 757 bp in length
*   16194    16273: gap of unknown length
*   16274    17087: contig of 814 bp in length
*   17088    17167: gap of unknown length
*   17168    18800: contig of 1633 bp in length
*   18801    18880: gap of unknown length
*   18881    19862: contig of 982 bp in length
*   19863    19942: gap of unknown length
*   19943    21322: contig of 1380 bp in length
*   21323    21402: gap of unknown length
*   21403    22651: contig of 1249 bp in length
*   22652    22731: gap of unknown length
*   22732    23794: contig of 1063 bp in length
*   23795    23874: gap of unknown length
*   23875    25796: contig of 1922 bp in length
*   25797    25876: gap of unknown length
*   25877    27307: contig of 1431 bp in length
*   27308    27387: gap of unknown length
*   27388    29595: contig of 2208 bp in length
*   29596    29675: gap of unknown length
*   29676    32153: contig of 2478 bp in length
*   32154    32233: gap of unknown length
*   32234    34285: contig of 2052 bp in length
*   34286    34365: gap of unknown length
*   34366    35868: contig of 1503 bp in length
*   35869    35948: gap of unknown length
*   35949    37609: contig of 1661 bp in length
*   37610    37689: gap of unknown length
*   37690    39593: contig of 1904 bp in length
*   39594    39673: gap of unknown length
*   39674    42093: contig of 2420 bp in length
*   42094    42173: gap of unknown length
*   42174    44407: contig of 2234 bp in length
*   44408    44487: gap of unknown length
*   44488    47045: contig of 2558 bp in length
*   47046    47125: gap of unknown length
*   47126    49672: contig of 2547 bp in length
*   49673    49752: gap of unknown length
*   49753    51251: contig of 1499 bp in length
*   51252    51331: gap of unknown length
*   51332    52547: contig of 1216 bp in length
*   52548    52627: gap of unknown length
*   52628    54821: contig of 2194 bp in length
*   54822    54901: gap of unknown length
*   54902    56540: contig of 1639 bp in length
*   56541    56620: gap of unknown length
*   56621    59274: contig of 2654 bp in length
*   59275    59354: gap of unknown length
*   59355    62336: contig of 2982 bp in length
*   62337    62416: gap of unknown length
*   62417    66304: contig of 3888 bp in length
*   66305    66384: gap of unknown length
*   66385    68911: contig of 2527 bp in length
*   68912    68991: gap of unknown length
*   68992    71844: contig of 2853 bp in length
*   71845    71924: gap of unknown length
*   71925    74959: contig of 3035 bp in length
*   74960    75039: gap of unknown length
*   75040    77200: contig of 2161 bp in length
*   77201    77280: gap of unknown length
*   77281    80390: contig of 3110 bp in length
*   80391    80470: gap of unknown length
*   80471    84267: contig of 3797 bp in length
*   84268    84347: gap of unknown length
*   84348    89197: contig of 4850 bp in length
*   89198    89277: gap of unknown length
*   89278    96088: contig of 6811 bp in length
*   96089    96168: gap of unknown length
*   96169   102457: contig of 6289 bp in length
*  102458   102537: gap of unknown length
*  102538   107506: contig of 4969 bp in length
*  107507   107586: gap of unknown length
*  107587   113433: contig of 5847 bp in length
*  113434   113513: gap of unknown length
*  113514   120777: contig of 7264 bp in length
*  120778   120857: gap of unknown length
*  120858   134488: contig of 13631 bp in length
*  134489   134568: gap of unknown length
*  134569   146676: contig of 12108 bp in length
*  146677   146756: gap of unknown length
*  146757   147420: contig of 664 bp in length
*  147421   147500: gap of unknown length
*  147501   148009: contig of 509 bp in length
*  148010   148089: gap of unknown length
*  148090   148420: contig of 331 bp in length
*  148421   148500: gap of unknown length
*  148501   149336: contig of 836 bp in length
*  149337   149416: gap of unknown length
*  149417   149628: contig of 212 bp in length
*  149629   149708: gap of unknown length
*  149709   150268: contig of 560 bp in length
*  150269   150348: gap of unknown length
*  150349   150645: contig of 297 bp in length
*  150646   150725: gap of unknown length
*  150726   151437: contig of 712 bp in length
*  151438   151517: gap of unknown length
*  151518   153103: contig of 1586 bp in length
*  153104   153183: gap of unknown length
*  153184   153649: contig of 466 bp in length
*  153650   153729: gap of unknown length
*  153730   154264: contig of 535 bp in length
*  154265   154344: gap of unknown length
*  154345   154946: contig of 602 bp in length
*  154947   155026: gap of unknown length
*  155027   155641: contig of 615 bp in length
*  155642   155721: gap of unknown length
*  155722   156401: contig of 680 bp in length
```

```
          *   156402   156481: gap of unknown length
          *   156482   156994: contig of 513 bp in length
          *   156995   157074: gap of unknown length
          *   157075   157625: contig of 551 bp in length
          *   157626   157705: gap of unknown length

alignment_scores:
           Quality:  132.00            Length:    232
             Ratio:  1.347               Gaps:      7
 Percent Similarity:  42.241   Percent Identity: 21.983

alignment_block:
 US-09-060-609-2 x AC011069   ..

 Align seg 1/1  to: AC011069  from: 1  to: 168266

      44 LeuProLeuLeuGlyGlyGlyGlySerGlySerGlyGluLysValSerVa 60
         :::|||:::   :::      |||||||||   |||   :::   |||
  116431 GTGCCCGTTCAGGCGCAACTGGGCTCCGGCATGGGACCCTCCTCATCGTC 116480

      60 lSerLysMetAlaAlaAla.........................TrpP 68
          |||:::   ::::::|||                           :::|
  116481 ATCTTCGGCATCTTCCGCCTCGGGCGGAGCTGGAAACTCGGCCTTCTATC 116530

      68 roSerGlyProSerAlaProGluAlaValThrAlaArgLeuValGlyVal 84
         ||   ||||||
  116531 CACTGGGACCG..................................... 116541

      85 LeuTrpPheValSerValThrThrGlyProTrpGlyAlaValAlaThrSe 101

  116541 ..................................................... 116541

     101 rAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLysValGlyGlnT 118
                       ::::::   |||:::   |||
  116542 ...............AATGTGATGTGCTCGTTTCTGCCGCGCGACTTTC 116575

     118 yrIleCysLysAspPro............................... 123
              ||||||||||||
  116576 TCGATTGCAAGGATCCCGTCGATCATCGGGAGAACGCCACGGCGCAGCAG 116625

     124 .....................LysIleAsnAspAlaThrGlnGluProVa 133
                              |||::::::   :::|||   |||   ||
  116626 GAAAAGAAATACGGATGCCTGAAGTTCGGTGGTTCCACGTACGAAGAGGT 116675

     133 lAsn..............CysThrAsnTyrThrAlaHisValSerCysP 145
         |:::              ||||||   :::   |||   ::::::|||:
  116676 GGAGCACGCCATGGTCTGGTGCACCGTGTTT...GCCGACATCGAGTGCT 116722

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
         ::                                |||||||:::|||
  116723 AC.............................GGTAATCGCACCTTT 116739

     162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
         :::   ::::::                        :::   |||   :::
  116740 CTGCGCGCCGGA.....................GTGCCCTGTGTCCGCTA 116768

     178 lAsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyT 195
          :::   :::   :::   :::::::::      |||:::::||||||:
  116769 CACGGATCACTACTTTGTGACCACACTGATCTACAGCATGCTGCTGGGTT 116818

     195 rpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeu 211
         ::||||||   |||||||||   ||||||      :::   ::::::
  116819 TCCTCGGTATGGATCGCTTCTGTCTCGGTCAAACGGGCACGGCTGTGGGC 116868

     212 LysPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIl 228
         |||:::   |||:::|||      |||:::      :::||||||   ||
  116869 AAACTGCTTACCATGGGCGGCGTGGGCGTTTGGTGGATCATCGACGTCAT 116918

     228 eLeuIleSerMetGlnIleValGlyProSerAspGlySerSerTyr 243
         ||||:::      :::   :::   |||:::|||||||||::::::
  116919 CCTCCTGATCACCAACAATTTGCTGCCCGAGGACGGCAGCAATTGG 116964

seq_name: gb_ba2:AE001582
```

```
seq_documentation_block:
LOCUS       AE001582    18753 bp    DNA             BCT       16-NOV-1999
DEFINITION  Borrelia burgdorferi plasmid lp21, complete plasmid sequence.
ACCESSION   AE001582
VERSION     AE001582.1  GI:6382381
KEYWORDS    .
SOURCE      Lyme disease spirochete.
  ORGANISM  Borrelia burgdorferi
            Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1  (bases 1 to 18753)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
            White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
            van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
  TITLE     A bacterial genome in flux: The twelve linear and nine circular
            extrachromosomal DNAs in an infectious isolate of the Lyme disease
            spirochete Borrelia burgdorferi
  JOURNAL   Mol. Microbiol. (1999) In press
REFERENCE   2  (bases 1 to 18753)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
            White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
            van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .18753
                     /organism="Borrelia burgdorferi"
                     /db_xref="taxon:139"
                     /plasmid="lp21"
     gene            184. .618
                     /gene="BBU01"
     CDS             184. .618
                     /gene="BBU01"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07699.1"
                     /db_xref="GI:6382388"
                     /translation="MENYQYFFNLFFIFKIIFIYLYIMYKSVKEQQEKGIDHTCRILI
                     LTETIFEINLILENYSQKTLLKKYNENLKNKNLPPSNISTMKKYLNQLEKEIKIIAKF
                     YFKNDQSLIYYKLNYTLEKIWLKLIELFYKELKQFIQKNTTT"
     gene            746. .1114
                     /gene="BBU02"
     CDS             746. .1114
                     /gene="BBU02"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07700.1"
                     /db_xref="GI:6382389"
                     /translation="MDGVINDTLVARMKKQIKFNKNKLIILVKTLDHMNKELLYSANK
                     TYNYVLIQNNFNEALAKTYQLRVNYKTLLEYLEILEKNPKVILKRPTNKENESFIGLY
                     TLLYPLEDCCTKIYNSHPNI"
     gene            complement(1238. .1357)
                     /gene="BBU03"
     CDS             complement(1238. .1357)
                     /gene="BBU03"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07701.1"
                     /db_xref="GI:6382390"
                     /translation="MVYYKLVRLYARIDVIISNEVGLREISLKEIVLSPIHLV"
     gene            1486. .2628
                     /gene="BBU04"
```

```
CDS             1486. .2628
                /gene="BBU04"
                /note="similar to PID:1174341 PID:1174349 GB:AE001584
                percent identity: 81.23; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF07693.1"
                /db_xref="GI:6382382"
                /translation="MLISFVKESLPMHKIKTTNNNPRNCYNKVQYKLIVLISTICYLN
                KTHKKYTQKTILYYFNENLRKNGQTISTLRTMQKYIYRLQKEIKVTTNYYQHMGVNSG
                TEIYYKLNYPKKDCYHKINQHFKEKKETRFQNRVTNYFNKNSDSKMGSVQCESCNSNK
                NNIKEERKINEIEKYQVINYFNKCNFSCKEILSILLNLNVDKDTMIKIIKTIKRTDIK
                AKNKNIYFPKSCSKEKQEKLKKILCNTQKELEKSGYNSEQLETNFQKIYENYKYKPHF
                IIENHKYSDLSYIKRKLEKSIERKKENSKQDYKNLRTNIFNILIEQLKKETNIEILKP
                IIKEYLNNQKKIEYKKVFRIYYSELLEIINGKHYSNLKKFRKKSVG"
gene            2868. .3656
                /gene="BBU05"
CDS             2868. .3656
                /gene="BBU05"
                /note="similar to GB:D26185 SP:P37522 PID:467381
                PID:580906 GB:AL009126 percent identity: 55.98; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="plasmid partition protein, putative"
                /protein_id="AAF07694.1"
                /db_xref="GI:6382383"
                /translation="MNKKPNILTIASLKGGVGKSTLTILFSYLLKDLGKKILIVDMDP
                QNSITSYFSNYVHNVEIYNTYSMLKGGVSFNNCVGKINDHIFIIPAHPILENFNLEPI
                DYKEIVLELRIEQNISAYEFDYILLDTPPHRDFLVKNALNITDHIIIPVQVERWSIES
                FTILMNMIENFQIIKKKIYTTSIVENQFIKNRNTLKDIERLLYEKYGSYVKGKIHFSN
                GIKVFINELIEPSSEESYYKEAKKTLHNILGYINHKNKYFIKNK"
gene            14633. .15235
                /gene="BBU06"
CDS             14633. .15235
                /gene="BBU06"
                /note="similar to GB:U03641 PID:458218 percent identity:
                61.64; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF07695.1"
                /db_xref="GI:6382384"
                /translation="MEDIYMSKKNKKKLFLNDRIEHNIIDNLSNLDTNLDKNLMVYKD
                LKEQLKLNLSNEIDTKIQRMRILYEIKTRKLYKYDGFKSFSQYLKTFVVAKTQAFFYL
                KLYSKILEGSISIDTIKELGFENIKKHIFKKKSFDFNNNNKSKLDNKNISIRIFMRDK
                ELYEFVKQDIKRASYIFKELFQNRKDVLTDIVTKYDNERN"
gene            15349. .15813
                /gene="BBU07"
CDS             15349. .15813
                /gene="BBU07"
                /note="similar to PID:1174341 PID:1174349 GB:AE001584
                percent identity: 81.71; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF07696.1"
                /db_xref="GI:6382385"
                /translation="MIEILKTIKRTEIKAKNKNIHFTKSCSKEKQEKLKEILCNTQKE
                LEKSGCNSEQLETNFQKIYENYKYKPHFIIENHKYSDLSYIKRKLEKSIEIKKENPQK
                DYESLKINIFHIFIEQLKKEINIETLKPLVKEYLNNQKKIKYTKVFDTYYTR"
gene            15791. .16084
                /gene="BBU08"
CDS             15791. .16084
                /gene="BBU08"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF07702.1"
                /db_xref="GI:6382391"
                /translation="MIHITRGDNIKKYLNSHTKVLITKDTFKSYFFRFLWDKKNIVIF
                DILKVYAQNLYKELINSEKTYNQVVVEIVENKELGYSINHRIIKVSSSKSIYY"
gene            complement(16228. .16548)
                /gene="BBU09"
CDS             complement(16228. .16548)
                /gene="BBU09"
                /note="similar to GB:U03641 PID:458212 percent identity:
                66.22; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF07697.1"
                /db_xref="GI:6382386"
                /translation="MGIFYFQKECKTMSKAVDEIYCHSCGKTIKKDAEICISCGVRNK
                QTENYNKLVVFLLCLFLGYLGVHRFYVGKIGTGMLYLFTFGFLYVGALIDLIRIATNK
                FECK"
gene            16603. .16800
                /gene="BBU10"
CDS             16603. .16800
                /gene="BBU10"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF07703.1"
                /db_xref="GI:6382392"
                /translation="MVFLIEQLKKETNIEILKPIIKEYLNNQKKIKYNKIFGIYYLEL
                LEIIKNRKHDLNLEKFSKKVV"
gene            16886. .17878
                /gene="BBU11"
CDS             16886. .17878
                /gene="BBU11"
                /note="similar to SP:Q10618 PID:1322425 GB:AL123456
                percent identity: 53.61; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF07698.1"
                /db_xref="GI:6382387"
                /translation="MIKTPIISEDQFKIALKLIKMGEIIVFPTDTVYAIGANIYDEYA
                VRMINLVKKKSINEPLMLYVDTIEKIKELSSHVPKSAKILMEKFSPGPLTYVLKKSIK
                IPKFINNNLDTVAIRIPKNKTALNLIRKLKNPLVVSSANLSKRPSSTSFSMALKELNG
                LVGGIIMPKKNKDSNIGIESTIIEFDSKDNVIILRPGAITKKMIEKVLGSRYKVNYAK
                ETKIDLKKMFSYRLRIPVHTFKRGDNIRKYLKTNTKVLITRDTFKFYFFSFLWDKKNI
                ILFDSLIEYAENLYKELVNAEEKYDQVIVEIVENEELGYAINNRIVKASSNNFM"
gene            17918. .18365
                /gene="BBU12"
                /note="This region contains an authentic frame shift and
                is not the result of a sequencing artifact; similar to
                GB:L31615 GB:L31418 PID:520783 PID:551742 percent
                identity: 61.90; identified by sequence similarity;
                putative"

alignment_scores:
           Quality:  117.00            Length:     63
             Ratio:  2.659               Gaps:      1
 Percent Similarity:  69.841   Percent Identity: 42.857

alignment_block:
 US-09-060-609-2 x AE001582/rev   ..

 Align seg 1/1  to reverse of: AE001582  from: 1  to: 18753

     173 IleSerCys.....AsgAsnValAsnGlyTyrSerTyrLysVa 185
         |||||||||           ::::::::::::::|||:::     ||
   16440 ATTTCTTGCGGGGTCAGAAATAAACAAACCGAAAACTACAATAAACTTGT 16391

     185 lAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheT 202
         |:::      |||   ||||||||||||:::|||||||:::   |||||||
```

```
  16390 GGTATTTTTACTATGCTTATTTCTTGGTTATTTAGGAGTTCACAGATTTT 16341

    202 yrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPhe 218
        ||:::|||      :::   |||:::|||   :::   |||   ||||||
  16340 ATGTAGGTAAAATAGGAACTGGTATGTTATACCTATTTACATTTGGATTT 16291

    219 CysGlyIleGlySerLeuIleAspPheIleLeuIleSer 231
          :::|||:::|||||||||:::|||   |||:::
  16290 TTATATGTTGGAGCTTTAATCGATCTTATTAGAATAGCA 16252

seq_name: gb_pr2:HSAC001228

seq_documentation_block:
LOCUS       HSAC001228 244254 bp    DNA             PRI       14-APR-1997
DEFINITION  244Kb Contig from Human Chromsome 11p15.5 spanning D11S1 through
            D11S25, complete sequence.
ACCESSION   AC001228 U90582
VERSION     AC001228.1  GI:1935053
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 244254)
  AUTHORS   Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D.,
            Davie,J., Davies,C.J., Davis,C., English,C., Fondon,T.,
            Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O.,
            Hahner,L., Harris,J., Higgins,M., Hinson,S., Megarity,C.,
            Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P.,
            Probst,S., Rayner,S., Reid,L., Schageman,J., Schilling,P.,
            Schultz,R., Shows,T., Syed,M., Valenzuela,D., Ward,T., Weissman,B.
            and Wilson,R.
  TITLE     Template
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244254)
  AUTHORS   Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D.,
            Davie,J., Davies,C.J., Davis,C., English,C., Fondon,T.,
            Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O.,
            Hahner,L., Harris,J., Higgins,M., Hinson,S., Megarity,C.,
            Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P.,
            Probst,S., Rayner,S., Reid,L., Schageman,J., Schilling,P.,
            Schultz,R., Shows,T., Syed,M., Valenzuela,D., Ward,T., Weissman,B.
            and Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-1997) Genome Science and Technology Center,
            University of Texas Southwestern Medical Center at Dallas, 5323
            Harry Hines Blvd, Dallas, TX 75235-8591, USA
FEATURES             Location/Qualifiers
     source          1. .244254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
BASE COUNT    59733 a  66956 c  64429 g  53105 t     31 others
ORIGIN


alignment_scores:
           Quality:  111.00            Length:    260
             Ratio:  0.949               Gaps:     17
 Percent Similarity:  45.000   Percent Identity: 24.231

alignment_block:
 US-09-060-609-2 x HSAC001228/rev   ..

 Align seg 1/1  to reverse of: HSAC001228  from: 1  to: 244254

      39 LeuLeuProPheSerLeuProLeuLeu.........GlyGlyGlyGlySe 52
         |||:::|||   |||   |||                |||||||||||::
  106458 CTCATGCCCAGTTCTCCTCCCCAGAGCCCAGGACCAGGGGGTGGGGGTGA 106409

      52 rGlySerGlyGluLysValSerValSerLysMetAla............A 65
         :||||||||||:::   |||   |||   :::::::::
  106408 GGGGTCGGGAAGTGCTGTG...GTCAGGAGGCTGGGAAGGCAACGGTACA 106362

      65 laAlaTrpProSer..............GlyProSerAlaProGluAla 76
           |||||||||:::              ||||||   :::|||:::
  106361 TGGCCTGGCCAGCAAATCAGCTATAGAAGGGACCCCCAGTTCCCAGGCCA 106312

      77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGl 93
                                     |||          :::   |||||
  106311 ..........................TGG.........ATGGGTACTGG 106298

      93 yPro..............TrpGlyAla..................... 97
         ||||            |||:::|||
  106297 TCCATAGCCTGTTAGTAGCTGGGCCGCACAGCAGGAGGTGAGCAGCGGGG 106248

      98 .........................ValAlaThrSerAlaGlyGlyGlu 105
                                  :::|||   ||||||:::::
  106247 CAAGCATGACCGCCTGAATTCCACCTCTCGCCAGATCAGCGGCAGCAGTA 106198

     106 GluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCysLysAs 122
         :::|||   :::   |||:::|||
  106197 GATTCTCACAGGGACGAGAACCTT..................... 106174

     122 pProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsnTyrT 139
                                        |||||||||:::::
  106173 .............................GTTGTGAACTGTGCACAAGAGG 106152

     139 hrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSer 155
         :::::::::|||      |||      |||:::         |||
  106151 GATCAAGGTTGAGCACT...CCCCATGAGAATCTAATACCAGATGATCTG 106105

     156 SerGlyAsnGlu............ThrHisPheThrGlyAsnGluValGl 168
         :::   |||:::           ||||||:::   |||   ::::::::
  106104 GGGTGAAACAGTTTCATCTGGCCCACCCACATCCGTGGA...AAAATTGC 106058

     168 yPhePheLys............................ProIleSerC 175
         :|||   |||                           |||:::||||
  106057 CTTCCATAAAACTGGTCCCTGGTGCGGAAAATGTTGGGGGCCACTGAGCT 106008

     175 ys..............................ArgAsnValAsnGlyTyr 181
         ||                             |||:::|||   |||:::
  106007 GTAGGATGAAGAGGCTGAGATAAGGGGACACCAGAGGAGTTCTGGGGTTC 105958

     182 SerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAl 198
         :::   ::::::   :::      |||   :::||||||||||::::
  105957 ATGTGGCTGGCAGTGTCTGGTGCGGTTCTGGGGTTCGGGTGGCTGGCAGT 105908

     198 aAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysT 215
         ::::         ||||||:::      |||:::   ::::::      :
  105907 GTCTGGTGCAGTTCTGGGATTCGGGTGGCTGGCA...GTGTCTGGTACGG 105861

     215 hrValGlyPheCysGlyIle.........GlySerLeuIleAspPhe... 227
         :::::||||||      :::         |||:::::::::   |||
  105860 TTCTGGGGTTCGGGTGGCTGGCAGTGTCTGGTACGGTTCTGGGGTTCGGG 105811

     228 ...IleLeuIleSerMetGlnIleValGly 236
         :::   :::|||      ::::::|||
  105810 TGGCTGGCAGTGTCTGGTACGGTTCTGGGG 105781

seq_name: gb_pr3:AC005950

seq_documentation_block:
LOCUS       AC005950   137932 bp    DNA             PRI       07-NOV-1998
DEFINITION  Homo sapiens Chromsome 11p15.5 PAC clone pDJ754h15 containing
            cdk-inhibitor p57/KIP2 (CDKN1C) gene, complete sequence.
ACCESSION   AC005950
VERSION     AC005950.1  GI:3850601
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137932)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
```

```
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 137932)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Butler,C., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1998) Genome Science & Technology Center,
            University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
COMMENT     IMPORTANT: This submission contains the entire insert of clone
            pDJ754h15.  pDJ754h15 comes from the RPCI-3 PAC library constructed
            at the Roswell Park Cancer Institute by the Pieter de Jong group.
            This clone has been finished according to strict quality criteria
            and attempts have been made to resolve all base calling problems
            such as compressions and repetitive elements.  The expected
            Phred/Phrap calculated errors/10kb is 0.97.  In addition, attempts
            have been made to assure over 99% of consensus base calls consist
            of either double-stranded coverage or 2 types of labeling chemistry
            on one strand.
            CHROMOSOMAL LOCUS: This PAC clone comes from the Chromosome 11p15.5
            Wilms tumor type 2 (WT2) region between ESTs CTSD and IGF2(INSL2).
            This region spanning approximately 2 Mbp is mapped between D11S2701
            and D11S1145.
FEATURES             Location/Qualifiers
     source          1. .137932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     repeat_region   25473. .25856
                     /rpt_family="MLT1"
     repeat_region   complement(43535. .43830)
                     /rpt_family="Tigger2"
     repeat_region   43844. .43896
                     /rpt_family="MER8"
     repeat_region   complement(46444. .46507)
                     /rpt_family="MIR"
     repeat_region   complement(55646. .55937)
                     /rpt_family="Alu"
     repeat_region   complement(59689. .59969)
                     /rpt_family="Alu"
     repeat_region   67064. .67345
                     /rpt_family="Alu"
     repeat_region   68833. .69223
                     /rpt_family="Alu"
     repeat_region   complement(75855. .76169)
                     /rpt_family="Alu"
     repeat_region   complement(76454. .76817)
                     /rpt_family="Alu"
     repeat_region   complement(77069. .77357)
                     /rpt_family="Alu"
     repeat_region   complement(77379. .77639)
                     /rpt_family="Alu"
     repeat_region   complement(79153. .79273)
                     /rpt_family="Alu"
     repeat_region   complement(79942. .80199)
                     /rpt_family="Alu"
     repeat_region   complement(80724. .81023)
                     /rpt_family="Alu"
     repeat_region   complement(81464. .81750)
                     /rpt_family="Alu"
     repeat_region   complement(81767. .82060)
                     /rpt_family="Alu"
     repeat_region   92346. .92627
                     /rpt_family="Alu"
     repeat_region   complement(93711. .93815)
                     /rpt_family="MIR"
     repeat_region   96725. .96791
                     /rpt_family="MIR"
     repeat_region   complement(97069. .97342)
                     /rpt_family="Alu"
     repeat_region   complement(97984. .98248)
                     /rpt_family="Alu"
     repeat_region   99165. .99370
                     /rpt_family="Alu"
     repeat_region   103948. .104243
                     /rpt_family="Alu"
     repeat_region   104381. .104461
                     /rpt_family="MER21"
     repeat_region   complement(108932. .109194)
                     /rpt_family="L1"
     repeat_region   109598. .110431
                     /rpt_family="L1"
     repeat_region   complement(122253. .122502)
                     /rpt_family="Alu"
     repeat_region   124475. .124783
                     /rpt_family="Alu"
     repeat_region   129005. .129292
                     /rpt_family="Alu"
     repeat_region   complement(129977. .130031)
                     /rpt_family="MER1"
     repeat_region   complement(130063. .130249)
                     /rpt_family="MER1"
     repeat_region   134047. .134113
                     /rpt_family="MER45"
     repeat_region   137718. .137926
                     /rpt_family="Alu"
BASE COUNT    29530 a  40683 c  40406 g  27313 t
ORIGIN


alignment_scores:
           Quality:  111.00            Length:    260
             Ratio:  0.949               Gaps:     17
 Percent Similarity:  45.000   Percent Identity: 24.231

alignment_block:
 US-09-060-609-2 x AC005950/rev  ..

 Align seg 1/1  to reverse of: AC005950  from: 1  to: 137932

      39 LeuLeuProPheSerLeuProLeuLeu.....GlyGlyGlyGlySe 52
         |||:::|||   |||   |||             ||||||||||::
  130416 CTCATGCCCAGTTCTCCTCCCCAGAGCCCAGGACCAGGGGGTGGGGGTGA 130367

      52 rGlySerGlyGluLysValSerValSerLysMetAla.....A 65
         :|||||||||:::   |||   |||   :::::::::
  130366 GGGGTCGGGAAGTGCTGTG...GTCAGGAGGCTGGGAAGGCAACGGTACA 130320

      65 laAlaTrpProSer.....GlyProSerAlaProGluAla 76
           |||||||||:::          ||||||   :::|||:::
  130319 TGGCCTGGCCAGCAAATCAGCTATAGAAGGGACCCCCAGTTCCCAGGCCA 130270

      77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGl 93
                                    |||       :::   |||||
  130269 .....TGG.....ATGGGTACTGG 130256

      93 yPro.....TrpGlyAla..... 97
         ||||        |||:::|||
  130255 TCCATAGCCTGTTAGTAGCTGGGCCGCACAGCAGGAGGTGAGCAGCGGGG 130206

      98 .....ValAlaThrSerAlaGlyGlyGlu 105
                              :::|||   ||||||::::::
  130205 CAAGCATGACCGCCTGAATTCCACCTCTCGCCAGATCAGCGGCAGCAGTA 130156

     106 GluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCysLysAs 122
         :::|||   :::   |||:::|||
  130155 GATTCTCACAGGGACGAGAACCTT..... 130132

     122 pProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsnTyrT 139
```

```
                                        |||||||||:::::::
  130131 ..........................GTTGTGAACTGTGCACAAGAGG 130110

     139 hrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSer 155
         :::::::::|||      |||      |||:::         |||
  130109 GATCAAGGTTGAGCACT...CCCCATGAGAATCTAATACCAGATGATCTG 130063

     156 SerGlyAsnGlu............ThrHisPheThrGlyAsnGluValGl 168
         :::   |||:::           ||||||:::  |||   ::::::::
  130062 GGGTGAAACAGTTTCATCTGGCCCACCCACATCCGTGGA...AAAATTGC 130016

     168 yPhePheLys..............................ProIleSerC 175
         :|||   |||                             |||:::||||
  130015 CTTCCATAAAACTGGTCCCTGGTGCGGAAAATGTTGGGGGCCACTGAGCT 129966

     175 ys..............................ArgAsnValAsnGlyTyr 181
         ||                             |||:::|||   |||:::
  129965 GTAGGATGAAGAGGCTGAGATAAGGGGACACCAGAGGAGTTCTGGGGTTC 129916

     182 SerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAl 198
         :::   ::::::  :::      |||   :::||||||||:::::
  129915 ATGTGGCTGGCAGTGTCTGGTGCGGTTCTGGGGTTCGGGTGGCTGGCAGT 129866

     198 aAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysT 215
         ::::        ||||||:::      |||:::   ::::::      :
  129865 GTCTGGTGCAGTTCTGGGATTCGGGTGGCTGGCA...GTGTCTGGTACGG 129819

     215 hrValGlyPheCysGlyIle.........GlySerLeuIleAspPhe... 227
         :::::||||||      :::        |||::::::::::   |||
  129818 TTCTGGGGTTCGGGTGGCTGGCAGTGTCTGGTACGGTTCTGGGGTTCGGG 129769

     228 ...IleLeuIleSerMetGlnIleValGly 236
         :::   :::|||      ::::::|||
  129768 TGGCTGGCAGTGTCTGGTACGGTTCTGGGG 129739

seq_name: gb_htg7:AC013791

seq_documentation_block:
LOCUS       AC013791   180077 bp    DNA             HTG       09-DEC-1999
DEFINITION  Homo sapiens clone RP11-19N21, WORKING DRAFT SEQUENCE, 16 unordered
            pieces.
ACCESSION   AC013791
VERSION     AC013791.2  GI:6553989
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180077)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-19N21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180077)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 1999 this sequence version replaced gi:6425699.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3857
                Center clone name: 19_N_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 142771 bases at least Q40
                Consensus quality: 163160 bases at least Q30
                Consensus quality: 173452 bases at least Q20
                Insert size: 173000; agarose-fp
                Insert size: 180077; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1311: contig of 1311 bp in length
            *                   gap of unknown length
            *     1312     2589: contig of 1278 bp in length
            *                   gap of unknown length
            *     2590     3846: contig of 1257 bp in length
            *                   gap of unknown length
            *     3847     5399: contig of 1553 bp in length
            *                   gap of unknown length
            *     5400     6617: contig of 1218 bp in length
            *                   gap of unknown length
            *     6618     8121: contig of 1504 bp in length
            *                   gap of unknown length
            *     8122    11128: contig of 3007 bp in length
            *                   gap of unknown length
            *    11129    14747: contig of 3619 bp in length
            *                   gap of unknown length
            *    14748    20411: contig of 5664 bp in length
            *                   gap of unknown length
            *    20412    32982: contig of 12571 bp in length
            *                   gap of unknown length
            *    32983    45907: contig of 12925 bp in length
            *                   gap of unknown length
            *    45908    60419: contig of 14512 bp in length
            *                   gap of unknown length
            *    60420    75595: contig of 15176 bp in length
            *                   gap of unknown length
            *    75596    92605: contig of 17010 bp in length
            *                   gap of unknown length
            *    92606   122499: contig of 29894 bp in length
            *                   gap of unknown length
            *   122500   180077: contig of 57578 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180077
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-19N21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    38004 a  51475 c  50983 g  39602 t     13 others
ORIGIN


alignment_scores:
           Quality:  111.00            Length:    260
             Ratio:  0.949               Gaps:     17
```

```
 Percent Similarity:  45.000   Percent Identity: 24.231

alignment_block:
 US-09-060-609-2 x AC013791   ..

 Align seg 1/1  to: AC013791  from: 1  to: 180077

      39 LeuLeuProPheSerLeuProLeuLeu.........GlyGlyGlyGlySe 52
         |||:::|||   |||   |||                ||||||||||||:: 
   69387 CTCATGCCCAGTTCTCCTCCCCAGAGCCCAGGACCAGGGGGTGGGGGTGA 69436

      52 rGlySerGlyGluLysValSerValSerLysMetAla............A 65
         :|||||||||:::   |||   |||   :::::::::             
   69437 GGGGTCGGGAAGTGCTGTG...GTCAGGAGGCTGGGAAGGCAACGGTACA 69483

      65 laAlaTrpProSer..............GlyProSerAlaProGluAla 76
           ||||||||||:::               ||||||   :::|||:::    
   69484 TGGCCTGGCCAGCAAATCAGCTATAGAAGGGACCCCCAGTTCCCAGGCCA 69533

      77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGl 93
                                    |||          :::   |||||  
   69534 ..........................TGG.........ATGGGTACTGG 69547

      93 yPro..............TrpGlyAla..................... 97
         ||||               |||:::|||                    
   69548 TCCATAGCCTGTTAGTAGCTGGGCCGCACAGCAGGAGGTGAGCAGCGGGG 69597

      98 .........................ValAlaThrSerAlaGlyGlyGlu 105
                                  :::|||   ||||||::::::    
   69598 CAAGCATGACCGCCTGAATTCCACCTCTCGCCAGATCAGCGGCAGCAGTA 69647

     106 GluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCysLysAs 122
         :::|||   :::   |||:::|||                          
   69648 GATTCTCACAGGGACGAGAACCTT........................... 69671

     122 pProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsnTyrT 139
                                       |||||||||::::::    
   69672 ..........................GTTGTGAACTGTGCACAAGAGG 69693

     139 hrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSer 155
           :::::::::|||       |||       |||:::          |||  
   69694 GATCAAGGTTGAGCACT...CCCCATGAGAATCTAATACCAGATGATCTG 69740

     156 SerGlyAsnGlu............ThrHisPheThrGlyAsnGluValGl 168
         :::   |||:::             ||||||:::   |||   ::::::::  
   69741 GGGTGAAACAGTTTCATCTGGCCCACCCACATCCGTGGA...AAAATTGC 69787

     168 yPhePheLys............................ProIleSerC 175
         :|||   |||                            |||:::||||  
   69788 CTTCCATAAAACTGGTCCCTGGTGCGGAAAATGTTGGGGGCCACTGAGCT 69837

     175 ys..............................ArgAsnValAsnGlyTyr 181
         ||                               |||:::|||   |||:::  
   69838 GTAGGATGAAGAGGCTGAGATAAGGGGACACCAGAGGAGTTCTGGGGTTC 69887

     182 SerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAl 198
         :::   ::::::   :::       |||   :::||||||||||:::::  
   69888 ATGTGGCTGGCAGTGTCTGGTGCGGTTCTGGGGTTCGGGTGGCTGGCAGT 69937

     198 aAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysT 215
         ::::          |||||||:::       |||:::   ::::::      :  
   69938 GTCTGGTGCAGTTCTGGGATTCGGGTGGCTGGCA...GTGTCTGGTACGG 69984

     215 hrValGlyPheCysGlyIle.........GlySerLeuIleAspPhe... 227
         :::::||||||       :::          |||::::::::::   |||  
   69985 TTCTGGGGTTCGGGTGGCTGGCAGTGTCTGGTACGGTTCTGGGGTTCGGG 70034

     228 ...IleLeuIleSerMetGlnIleValGly 236
            :::   :::|||      :::::::|||  
   70035 TGGCTGGCAGTGTCTGGTACGGTTCTGGGG 70064

seq_name: gb_htg3:AC010844
```

```
seq_documentation_block:
LOCUS       AC010844    92509 bp    DNA             HTG       08-OCT-1999
DEFINITION  Drosophila melanogaster chromosome X clone BACR03F03 (D882) RPCI-98
            03.F.3 map 12C-12D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 100 unordered pieces.
ACCESSION   AC010844
VERSION     AC010844.4  GI:6016642
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 92509)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92509)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Oct 8, 1999 this sequence version replaced gi:6006066.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 100 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      664: contig of 664 bp in length
            *      665      744: gap of unknown length
            *      745     1298: contig of 554 bp in length
            *     1299     1378: gap of unknown length
            *     1379     1987: contig of 609 bp in length
            *     1988     2067: gap of unknown length
            *     2068     2719: contig of 652 bp in length
            *     2720     2799: gap of unknown length
            *     2800     3932: contig of 1133 bp in length
            *     3933     4012: gap of unknown length
            *     4013     4712: contig of 700 bp in length
            *     4713     4792: gap of unknown length
            *     4793     5546: contig of 754 bp in length
            *     5547     5626: gap of unknown length
            *     5627     6428: contig of 802 bp in length
            *     6429     6508: gap of unknown length
            *     6509     7234: contig of 726 bp in length
            *     7235     7314: gap of unknown length
            *     7315     8210: contig of 896 bp in length
            *     8211     8290: gap of unknown length
            *     8291     8957: contig of 667 bp in length
```

```
*     8958     9037: gap of unknown length
*     9038     9940: contig of 903 bp in length
*     9941    10020: gap of unknown length
*    10021    10867: contig of 847 bp in length
*    10868    10947: gap of unknown length
*    10948    11592: contig of 645 bp in length
*    11593    11672: gap of unknown length
*    11673    12300: contig of 628 bp in length
*    12301    12380: gap of unknown length
*    12381    12785: contig of 405 bp in length
*    12786    12865: gap of unknown length
*    12866    13713: contig of 848 bp in length
*    13714    13793: gap of unknown length
*    13794    14731: contig of 938 bp in length
*    14732    14811: gap of unknown length
*    14812    15574: contig of 763 bp in length
*    15575    15654: gap of unknown length
*    15655    16762: contig of 1108 bp in length
*    16763    16842: gap of unknown length
*    16843    17720: contig of 878 bp in length
*    17721    17800: gap of unknown length
*    17801    18710: contig of 910 bp in length
*    18711    18790: gap of unknown length
*    18791    19193: contig of 403 bp in length
*    19194    19273: gap of unknown length
*    19274    20195: contig of 922 bp in length
*    20196    20275: gap of unknown length
*    20276    21358: contig of 1083 bp in length
*    21359    21438: gap of unknown length
*    21439    22656: contig of 1218 bp in length
*    22657    22736: gap of unknown length
*    22737    23743: contig of 1007 bp in length
*    23744    23823: gap of unknown length
*    23824    25367: contig of 1544 bp in length
*    25368    25447: gap of unknown length
*    25448    26594: contig of 1147 bp in length
*    26595    26674: gap of unknown length
*    26675    27986: contig of 1312 bp in length
*    27987    28066: gap of unknown length
*    28067    28982: contig of 916 bp in length
*    28983    29062: gap of unknown length
*    29063    29924: contig of 862 bp in length
*    29925    30004: gap of unknown length
*    30005    30968: contig of 964 bp in length
*    30969    31048: gap of unknown length
*    31049    32547: contig of 1499 bp in length
*    32548    32627: gap of unknown length
*    32628    33808: contig of 1181 bp in length
*    33809    33888: gap of unknown length
*    33889    34986: contig of 1098 bp in length
*    34987    35066: gap of unknown length
*    35067    35803: contig of 737 bp in length
*    35804    35883: gap of unknown length
*    35884    36294: contig of 411 bp in length
*    36295    36374: gap of unknown length
*    36375    37470: contig of 1096 bp in length
*    37471    37550: gap of unknown length
*    37551    39563: contig of 2013 bp in length
*    39564    39643: gap of unknown length
*    39644    41050: contig of 1407 bp in length
*    41051    41130: gap of unknown length
*    41131    42389: contig of 1259 bp in length
*    42390    42469: gap of unknown length
*    42470    44106: contig of 1637 bp in length
*    44107    44186: gap of unknown length
*    44187    45257: contig of 1071 bp in length
*    45258    45337: gap of unknown length
*    45338    46808: contig of 1471 bp in length
*    46809    46888: gap of unknown length
*    46889    49104: contig of 2216 bp in length
*    49105    49184: gap of unknown length
*    49185    50451: contig of 1267 bp in length
*    50452    50531: gap of unknown length
*    50532    52592: contig of 2061 bp in length
*    52593    52672: gap of unknown length
*    52673    55002: contig of 2330 bp in length
*    55003    55082: gap of unknown length
*    55083    57023: contig of 1941 bp in length
*    57024    57103: gap of unknown length
*    57104    59408: contig of 2305 bp in length
*    59409    59488: gap of unknown length
*    59489    60015: contig of 527 bp in length
*    60016    60095: gap of unknown length
*    60096    60452: contig of 357 bp in length
*    60453    60532: gap of unknown length
*    60533    61086: contig of 554 bp in length
*    61087    61166: gap of unknown length
*    61167    61961: contig of 795 bp in length
*    61962    62041: gap of unknown length
*    62042    62577: contig of 536 bp in length
*    62578    62657: gap of unknown length
*    62658    63205: contig of 548 bp in length
*    63206    63285: gap of unknown length
*    63286    63767: contig of 482 bp in length
*    63768    63847: gap of unknown length
*    63848    64432: contig of 585 bp in length
*    64433    64512: gap of unknown length
*    64513    65290: contig of 778 bp in length
*    65291    65370: gap of unknown length
*    65371    66186: contig of 816 bp in length
*    66187    66266: gap of unknown length
*    66267    66724: contig of 458 bp in length
*    66725    66804: gap of unknown length
*    66805    67422: contig of 618 bp in length
*    67423    67502: gap of unknown length
*    67503    68287: contig of 785 bp in length
*    68288    68367: gap of unknown length
*    68368    68994: contig of 627 bp in length
*    68995    69074: gap of unknown length
*    69075    69731: contig of 657 bp in length
*    69732    69811: gap of unknown length
*    69812    70509: contig of 698 bp in length
*    70510    70589: gap of unknown length
*    70590    71259: contig of 670 bp in length
*    71260    71339: gap of unknown length
*    71340    71975: contig of 636 bp in length
*    71976    72055: gap of unknown length
*    72056    72256: contig of 201 bp in length
*    72257    72336: gap of unknown length
*    72337    73239: contig of 903 bp in length
*    73240    73319: gap of unknown length
*    73320    73974: contig of 655 bp in length
*    73975    74054: gap of unknown length
*    74055    74743: contig of 689 bp in length
*    74744    74823: gap of unknown length
*    74824    75343: contig of 520 bp in length
*    75344    75423: gap of unknown length
*    75424    76021: contig of 598 bp in length

alignment_scores:
            Quality:  110.50            Length:     91
              Ratio:  1.973               Gaps:      1
 Percent Similarity:  61.538   Percent Identity: 29.670

alignment_block:
 US-09-060-609-2 x AC010844   ..

 Align seg 1/1  to: AC010844  from: 1  to: 92509

     153 LysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPh 169
         |||:::      ||||||:::|||   :::   ::::::
    2821 AAAGAGTGCTACGGTAATCGCACCTTTCTGCGCGCCGGA.......... 2859

     169 ePheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValA 186
                  :::   |||   :::   :::   :::   :::   ::::
    2860 .........GTGCCCTGTGTCCGCTACACGGATCACTACTTTGTGACCA 2899
```

```
 186 laValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyr 202
     :::::      |||::::::||||||:::||||||   |||||||||
2900 CACTGATCTACAGCATGCTGCTGGGTTTCCTCGGTATGGATCGCTTCTGT 2949

 203 LeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCy 219
     ||||||      :::   ::::::   |||:::   |||:::|||
2950 CTCGGTCAAACGGGCACGGCTGTGGGCAAACTGCTTACCATGGGCGGCGT 2999

 219 sGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleValG 236
      |||:::      :::|||||||   ||||||:::      :::   :::
3000 GGGCGTTTGGTGGATCATCGACGTCATCCTCCTGATCACCAACAATTTGC 3049

 236 lyProSerAspGlySerSerTyr 243
       |||:::||||||||||::::::
3050 TGCCCGAGGACGGCAGCAATTGG 3072
```

```
OM of: US-09-060-609-2 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 11, 2000  8:09 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09060609/runat_09022000_150439_25205/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09060609 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

[redacted]ch information block:
[redacted]ery: US-09-060-609-2
Query length: 269
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 50.600000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
N_Geneseq_36:X05735   + 1439.00 2529.46  2.7e-133  810  ! Human beta-amyloid peptide-bind
N_Geneseq_36:X41191   +  779.00 1368.07  1.3e-68   440  ! Human secreted protein 5' EST S
N_Geneseq_36:X41259   +  760.50 1334.97  9.3e-67   455  ! Human secreted protein 5' EST S
N_Geneseq_36:V40276   +  108.00  163.53  0.1654   2162  ! Human equilibrative nucleoside
N_Geneseq_36:V49582   +  108.00  165.98  0.1207   1721  ! Human fibrosarcoma cell line H
N_Geneseq_36:V49583   +  108.00  170.61  0.0667   1119  ! Human fibrosarcoma cell line H
N_Geneseq_36:X20370   +  105.00  170.12  0.0710   715  ! Borrelia burgdorferi polynuclec
N_Geneseq_36:V84557   +   96.00  142.47  2.46     2127  ! Human secreted protein gene 14
N_Geneseq_36:T39751   -   91.50  137.99  4.38     1539  ! Human mucosal addressin cell a
N_Geneseq_36:X35280   -   91.50  137.99  4.38     1539  ! cDNA encoding human MAdCAM-1.
N_Geneseq_36:V22918   +   88.00  127.03  17.85    2397  ! Human GlyT-2 transporter G1387
N_Geneseq_36:V22914   +   87.00  125.26  22.40    2397  ! Human GlyT-2 transporter T836A
N_Geneseq_36:V22906   +   87.00  125.26  22.40    2397  ! Human GlyT-2 variant gene sequ
N_Geneseq_36:V22900   +   87.00  129.63  12.79    1597  ! Fragment of the human GlyT-2 g
N_Geneseq_36:T39750   -   86.50  128.56  14.67    1624  ! Human mucosal addressin cell a
N_Geneseq_36:V80679   +   86.50  132.06  9.36     1173  ! Human DNAX toll-like receptor
N_Geneseq_36:V72100   -   86.50  117.50  60.56    4541  ! Human RON receptor cDNA. Treat
N_Geneseq_36:X35279   -   86.50  128.56  14.67    1624  ! cDNA encoding human MAdCAM-1.
N_Geneseq_36:N50491   -   86.00  123.97  26.44    2293  ! Sequence of herpes simplex vir
N_Geneseq_36:V22917   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter G1371
N_Geneseq_36:V22919   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter G1829
N_Geneseq_36:V22920   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter A1831
N_Geneseq_36:V22921   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter G2103
N_Geneseq_36:V22922   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter T2198
N_Geneseq_36:V22923   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter A2203
N_Geneseq_36:V22924   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter C342G
N_Geneseq_36:V22925   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter C352T
N_Geneseq_36:V22926   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter T733C
N_Geneseq_36:V22927   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter A777G
N_Geneseq_36:V22929   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter G951A
N_Geneseq_36:V22910   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter T6C m
N_Geneseq_36:V22911   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter G304A
N_Geneseq_36:V22912   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter C371T
N_Geneseq_36:V22913   +   86.00  123.49  28.11    2397  ! Human GlyT-2 transporter C571T
N_Geneseq_36:V22905   +   86.00  123.49  28.11    2397  ! Human GlyT-2 gene sequence. Hu
N_Geneseq_36:V22909   +   86.00  123.49  28.11    2397  ! DNA encoding a human GlyT-2 tr
N_Geneseq_36:V22932   +   86.00  123.49  28.11    2397  ! DNA encoding a human GlyT-2 tr
N_Geneseq_36:X24307   -   85.50  126.32  19.54    1696  ! Rat bone mineralisation protei
N_Geneseq_36:T42080   +   85.00  128.82  14.18    1238  ! Rat LAMP coding sequence. DNA
N_Geneseq_36:T42081   +   85.00  130.97  10.77    1014  ! Rat LAMP coding sequence. DNA
N_Geneseq_36:T42083   +   85.00  131.90  9.56      930  ! Rat mature LAMP coding sequence
N_Geneseq_36:V22916   +   85.00  121.72  35.27    2397  ! Human GlyT-2 transporter A1
N_Geneseq_36:V62176   -   84.50   79.01  8.4e+03  117213  ! HSV-2 strain SB5 Contig I
N_Geneseq_36:X04125   +   84.50  120.46  41.47    2483  ! Human Factor B encoding cDN
N_Geneseq_36:Q20349   +   84.00  120.93  39.05    2189  ! Thromboplastin clone 2b-Apr

seq_name: N_Geneseq_36:X05735

seq_documentation_block:
ID   X05735 standard; mRNA; 810 BP.
AC   X05735;
DT   27-APR-1999  (first entry)
DE   Human beta-amyloid peptide-binding protein (BBP) encoding mRNA.
KW   Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;
KW   human; Alzheimer's disease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .810
FT                   /*tag= a
FT                   /product= "BBP"
PN   WO9846636-A2.
PD   22-OCT-1998.
PF   14-APR-1998; U07462.
PR   16-APR-1997; US-064583.
PA   (AMHP ) AMERICAN HOME PROD CORP.
PI   Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,
PI   Walker SG;
DR   WPI; 99-080736/07.
DR   P-PSDB; W94291.
PT   Polynucleotide encoding beta-amyloid peptide binding protein - used
PT   to identify inhibitors of beta-amyloid peptide for treating
PT   Alzheimer's disease
PS   Claim 1; Pages 43-44; 59pp; English.
CC   This represents a nucleotide sequence encoding a beta-amyloid peptide
CC   binding protein (BBP). The polynucleotide comprising the entire BBP
CC   nucleotide sequence of clone BBP1-fl is deposited under the accession
CC   number ATCC 98617. The polynucleotide comprising a fragment of BBP
CC   (nucleotides 202-807 of the full length BBP) of clone pEK196 is deposited
CC   as ATCC 98399. Host cells transformed with a vector comprising the BBP
CC   nucleic acid are used for the recombinant production of the protein. The
CC   protein can be used in a method for diagnosing a disease characterised by
CC   aberrant expression of human beta-amyloid protein (BAP). The protein can
CC   also be used in a method for screening for compounds which regulate
CC   expression of a BAP binding protein. The proteins, antibodies and
CC   identified compounds can be used in the treatment or prevention of
CC   Alzheimer's disease.
SQ   Sequence   810 BP;    204 A;    183 C;    202 G;    221 T;


alignment_scores:
           Quality: 1439.00              Length:    269
             Ratio:  5.349                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-060-609-2 x X05735   ..

 Align seg 1/1  to: X05735  from: 1  to: 810

       1 MetHisIleLeuLysGlySerProAsnValIleProArgAlaHisGlyGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGCATATTTTAAAAGGGTCTCCCAATGTGATTCCACGGGCTCACGGGCA 50

      17 nLysAsnThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GAAGAACACGCGAAGAGACGGAACTGGCCTCTATCCTATGCGAGGTCCCT 100

      34 heLysAsnLeuAlaLeuLeuProPheSerLeuProLeuLeuGlyGlyGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 TTAAGAACCTCGCCCTGTTGCCCTTCTCCCTCCCGCTCCTGGGCGGAGGC 150

      51 GlySerGlySerGlyGluLysValSerValSerLysMetAlaAlaAlaTr 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAGATGGCGGCCGCCTG 200
```

```
 67 pProSerGlyProSerAlaProGluAlaValThrAlaArgLeuValGlyV 84
    |||||||||||||||||||||||||||||||||||||||||||||||||||
201 GCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTG 250

 84 alLeuTrpPheValSerValThrThrGlyProTrpGlyAlaValAlaThr 100
    |||||||||||||||||||||||||||||||||||||||||||||||||||
251 TCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300

101 SerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLysValGlyGl 117
    |||||||||||||||||||||||||||||||||||||||||||||||||||
301 TCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACA 350

117 nTyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluProValA 134
    |||||||||||||||||||||||||||||||||||||||||||||||||||
351 ATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTA 400

134 snCysThrAsnTyrThrAlaHisValSerCysPheProAlaProAsnIle 150
    |||||||||||||||||||||||||||||||||||||||||||||||||||
401 ACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATA 450

151 ThrCysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluVa 167
    |||||||||||||||||||||||||||||||||||||||||||||||||||
451 ACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGT 500

167 lGlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrL 184
    |||||||||||||||||||||||||||||||||||||||||||||||||||
501 TGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACA 550

184 ysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArg 200
    |||||||||||||||||||||||||||||||||||||||||||||||||||
551 AAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGA 600

201 PheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGl 217
    |||||||||||||||||||||||||||||||||||||||||||||||||||
601 TTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGG 650

217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
    |||||||||||||||||||||||||||||||||||||||||||||||||||
651 GTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGA 700

234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
    |||||||||||||||||||||||||||||||||||||||||||||||||||
701 TTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACC 750

251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
    |||||||||||||||||||||||||||||||||||||||||||||||||||
751 AGACTTACAAGACTGAGTATTACTAATGAAACATTTAGAAAAACGCAATT 800

267 uTyrPro 269
    |||||||
801 ATATCCA 807
```

seq_name: N_Geneseq_36:X41191

seq_documentation_block:

```
ID   X41191 standard; cDNA; 440 BP.
AC   X41191;
DT   17-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID NO:135.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   P-PSDB; Y12358.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 1; Page 315; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   440 BP;    107 A;    103 C;    114 G;    114 T;
```

alignment_scores:

```
           Quality:  779.00          Length:    146
             Ratio:  5.372             Gaps:      0
 Percent Similarity:  99.315   Percent Identity: 99.315
```

alignment_block:

US-09-060-609-2 x X41191 ..

Align seg 1/1 to: X41191 from: 1 to: 440

```
 56 GluLysValSerValSerLysMetAlaAlaAlaTrpProSerGlyProSe 72
    |||||||||||||||||||||||||||||||||||   ||||||||||||
  3 GAGAAAGTGTCGGTCTCCAAGATGGCGGCCGCCTGGCSDTCTGGTCCGTC 52

 72 rAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrpPheValS 89
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 53 TGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCT 102

 89 erValThrThrGlyProTrpGlyAlaValAlaThrSerAlaGlyGlyGlu 105
    |||||||||||||||||||||||||||||||||||||||||||||||||||
103 CAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAG 152

106 GluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCysLysAs 122
    |||||||||||||||||||||||||||||||||||||||||||||||||||
153 GAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGA 202

122 pProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsnTyrT 139
    |||||||||||||||||||||||||||||||||||||||||||||||||||
203 TCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACA 252

139 hrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSer 155
    |||||||||||||||||||||||||||||||||||||||||||||||||||
253 CAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCC 302

156 SerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheLysPr 172
    |||||||||||||||||||||||||||||||||||||||||||||||||||
303 AGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCC 352

172 oIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaL 189
    |||||||||||||||||||||||||||||||||||||||||||||||||||
353 CATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCAT 402

189 euSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPhe 201
    |||||||||||||||||||||||||||||||||||||
403 TGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTT 440
```

seq_name: N_Geneseq_36:X41259

```
seq_documentation_block:
ID   X41259 standard; cDNA; 455 BP.
AC   X41259;
DT   17-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID NO:203.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
     WPI; 99-153778/13.
     P-PSDB; Y12426.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 1; Page 456; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   455 BP;    102 A;    107 C;    115 G;    122 T;


alignment_scores:
           Quality:  760.50            Length:    152
             Ratio:  5.173               Gaps:      2
 Percent Similarity:  96.711   Percent Identity: 94.079

alignment_block:
 US-09-060-609-2 x X41259   ..

 Align seg 1/1  to: X41259  from: 1  to: 455

      60 ValSerLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAl 76
         ||||||||||||||||||||||||||||||||||:::|||||||||||
       3 GTCTCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGKCTGCTCCGGAGGC 52

      76 aValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrG 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAG 102

      93 lyProTrpGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLeuLys 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 GACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 152

     110 CysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAs 126
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAA 202

     126 nAspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValS 143
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     203 TGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTT 252

     143 erCysPheProAlaProAsnIleThrCysLysAspSerSerGlyAsnGlu 159
         ||||||||||||||||||||||||||||||||||:::|||||||||||
     253 CCTGTTTTCCAGCACCCAACATAACTTGTAAGGATNCCAGTGGCAATGAA 302

     160 ThrHisPheThrGlyAsnGluValGlyPhePheLysProIleSerCysAr 176
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 ACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCG 352

     176 gAsnValAsnGlyTyrSerTyrLysVal.AlaValAlaLeuSerLeuPhe 192
         ||||||||||||||||||||||      :::   :::   |||||||||
     353 AAATGTAAATGGCTATTCCTACNNTKAGCAGT...NNNWTGTCTCTTTTT 399

     193 LeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGl 209
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     400 CTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGG 449

     209 yLeu 210
         ||||
     450 TTTG 453

seq_name: N_Geneseq_36:V40276

seq_documentation_block:
ID   V40276 standard; cDNA; 2162 BP.
AC   V40276;
DT   13-OCT-1998  (first entry)
DE   Human equilibrative nucleoside transporter 1 encoding cDNA.
KW   Human; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;
KW   rENT2; coronary; cerebrovascular anoxia; viral infection; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             179. .1549
FT                   /*tag= a
FT                   /product= "equilibrative nucleoside transporter 1"
PN   WO9829437-A2.
PD   09-JUL-1998.
PF   30-DEC-1997; IB1657.
PR   03-NOV-1997; US-064004.
PR   30-DEC-1996; US-034083.
PA   (UYAL-) UNIV ALBERTA.
PA   (UYLE-) UNIV LEEDS.
PI   Baldwin SA, Cass CE, Young JD;
DR   WPI; 98-388035/33.
DR   P-PSDB; W69555.
PT   Newly isolated equilibrative nucleoside transporter protein(s) and
PT   gene(s) - used to develop products for treating disorder(s)
PT   associated with the transporter(s) and for use with nucleoside
PT   drug(s)
PS   Claim 21; Fig 2; 97pp; English.
CC   The present sequence encodes a substantially purified equilibrative
CC   nucleoside transporter (ENT), human ENT1 (hENT1). ENTs can transport a
CC   variety of purines and pyrimidines, including adenosine, uridine,
CC   guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are
CC   bidirectional, they transport a suitable permeant both into and out of
CC   cells. ENTs can be used as a tool for the development of new nucleoside
CC   drugs. Products from the present invention can be used for treating a
CC   subject having a disorder associated with an ENT. They can also be used
CC   with nucleoside drugs in the treatment of e.g. coronary or
CC   cerebrovascular anoxia, viral infection or cancer. The products (e.g.
CC   antibodies and oligonucleotides hybridising to nucleic acid sequences
CC   encoding ENTs) can also be used for detection and drug screening.
SQ   Sequence   2162 BP;    389 A;    638 C;    560 G;    575 T;


alignment_scores:
           Quality:  108.00            Length:    236
             Ratio:  0.947               Gaps:     14
 Percent Similarity:  48.305   Percent Identity: 23.305

alignment_block:
 US-09-060-609-2 x V40276   ..
```

```
Align seq 1/1  to: V40276  from: 1  to: 2162

     30 MetArgGlyProPheLysAsnLeuAla...LeuLeuPro......PheSe 43
        :::::::|||   :::   :::|||||||   |||||||||      :::::
    647 CTGCAGGGCAGCCTGTTTGGTCTGGCTGGCCTTCTGCCTGCCAGCTACAC 696

     43 rLeuProLeuLeuGlyGlyGlyGlySerGlySerGlyGluLysValSerV 60
        :   |||:::::::::|||   |||      :::|||      :::|||||
    697 GGCCCCCATCATGAGTGGCCAGGGCCTA...GCAGGCTTCTTTGCCTCCG 743

     60 alSerLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAla 76
        ||:::   |||   :::|||      ||||||   :::      |||:::
    744 TGGCC...ATGATCTGCGCTATTGCCAGTGGCTCGGAACTATCAGAAAGT 790

     77 ..............ValThrAlaArgLeuValGlyValLeuTrpPheVa 88
                       :::|||||||      |||   :::|||   :::::
    791 GCCTTCGGCTACTTTATCACAGCCTGTGCTGTTATCATTTTGACCATCAT 840

     88 lSerValThrThrGlyPro.....................TrpGlyAlaV 98
        :                |||                    :::      :
    841 CTGTTACCTGGGCCTGCCCCGCCTGGAATTCTACCGCTACTACCAGCAGC 890

     98 alAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLys 114
        ::      ::::::   ||||||::::::   |||   :::   ::::::
    891 TCAAGCTTGAAGGACCCGGGGAGCAGGAGACCAAGTTGGACCTCATTAGC 940

    115 ValGlyGlnTyrIleCysLysAspProLys..................... 124
           |||:::            :::|||:::
    941 AAAGGAGAG............GAGCCAAGAGCAGGCAAAGAGGAATCTGG 978

    125 .IleAsnAspAlaThrGlnGluProValAsn.................... 134
         ::::::   ::::::::::::|||:::|||
    979 AGTTTCAGTCTCCAACTCTCAGCCCACCAATGAAAGCCACTCTATCAAAG 1028

    135 .....................................CysThrAsnTyr 138
                                            |||      :::
   1029 CCATCCTGAAAAATATCTCAGTCCTGGCTTTCTCTGTCTGCTTCATCTTC 1078

    139 ThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSe 155
        |||      ::::::   |||||||||   ::::::      |||:::||
   1079 ACTATCACCATTGGGATGTTTCCAGCCGTGACTGTTGAGGTCAAGTCCAG 1128

    155 r...SerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheL 171
        |   :::|||::::::|||                  :::|||
   1129 CATCGCAGGCAGCAGCACCTGG..............GAACGTTACTTCA 1163

    171 ysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVal 187
          |||:::||||||                              :::
   1164 TTCCTGTGTCCTGT.....................TTCTTG 1183

    188 AlaLeuSerLeuPhe.LeuGlyTrpLeuGlyAlaAspArgPheTyr.... 202
        :::::::::::::|||    ||||||   ||||||:::::::::|||
   1184 ACTTTCAATATCTTTGACTGGTTGGGCCGGAGCCTCACAGCTGTATTCAT 1233

    203 .....LeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGly 217
             ||||||      |||   |||          :::|||      |||
   1234 GTGGCCTGGGAAGGACAGCCGCTGGCTGCCAAGCCTGGTGCTGGCCCGGC 1283

    218 PheCys 219
        :::|||
   1284 TGGTGT 1289

seq_name: N_Geneseq_36:V49582

seq_documentation_block:
ID   V49582 standard; cDNA to mRNA; 1721 BP.
AC   V49582;
DT   21-OCT-1998  (first entry)
DE   Human fibrosarcoma cell line HT-1080 clone HP10235 cDNA #1.
KW   Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW   differentiation; immune system; stimulator; suppressor; regulator;
KW   heamatopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW   haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             6. .1127
FT                   /*tag=  a
FT                   /product= "transmembrane domain containing protein"
PN   WO9821328-A2.
PD   22-MAY-1998.
PF   07-NOV-1997; J04056.
PR   13-NOV-1996; JP-301429.
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PI   Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR   WPI; 98-297932/26.
DR   P-PSDB; W64550.
PT   Human protein having transmembrane domain - useful for, e.g.
PT   research and nutrition
PS   Claim 4; Page 160-162; 205pp; English.
CC   V49550-V49599 are cDNA sequences which encode human proteins containing a
CC   transmembrane domain. These proteins can be used for, e.g. research and
CC   nutrition, and may have cytokine and cell proliferation/differentiation,
CC   immune stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC   activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC   receptor/ligand, anti-inflammatory or tumour inhibition activity.
SQ   Sequence   1721 BP;    291 A;    504 C;    447 G;    479 T;

alignment_scores:
            Quality:  108.00              Length:    236
              Ratio:  0.947                 Gaps:     14
 Percent Similarity:  48.305   Percent Identity: 23.305

alignment_block:
 US-09-060-609-2 x V49582   ..

 Align seq 1/1  to: V49582  from: 1  to: 1721

      30 MetArgGlyProPheLysAsnLeuAla...LeuLeuPro......PheSe 43
         ::::::|||   :::   :::||||||   |||||||||      :::::
     225 CTGCAGGGCAGCCTGTTTGGTCTGGCTGGCCTTCTGCCTGCCAGCTACAC 274

      43 rLeuProLeuLeuGlyGlyGlyGlySerGlySerGlyGluLysValSerV 60
         :   |||:::::::::|||   |||      :::|||      :::|||||
     275 GGCCCCCATCATGAGTGGCCAGGGCCTA...GCAGGCTTCTTTGCCTCCG 321

      60 alSerLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAla 76
         ||:::   |||   :::|||      ||||||   :::      |||:::
     322 TGGCC...ATGATCTGCGCTATTGCCAGTGGCTCGGAGCTATCAGAAAGT 368

      77 ..............ValThrAlaArgLeuValGlyValLeuTrpPheVa 88
                        :::|||||||      |||   :::|||   :::::
     369 GCCTTCGGCTACTTTATCACAGCCTGTGCTGTTATCATTTTGACCATCAT 418

      88 lSerValThrThrGlyPro.....................TrpGlyAlaV 98
         :                |||                    :::      :
     419 CTGTTACCTGGGCCTGCCCCGCCTGGAATTCTACCGCTACTACCAGCAGC 468

      98 alAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLys 114
         ::      ::::::   ||||||::::::   |||   :::   ::::::
     469 TCAAGCTTGAAGGACCCGGGGAGCAGGAGACCAAGTTGGACCTCATTAGC 518

     115 ValGlyGlnTyrIleCysLysAspProLys..................... 124
            |||:::            :::|||:::
     519 AAAGGAGAG............GAGCCAAGAGCAGGCAAAGAGGAATCTGG 556

     125 .IleAsnAspAlaThrGlnGluProValAsn.................... 134
          ::::::   ::::::::::::|||:::|||
     557 AGTTTCAGTCTCCAACTCTCAGCCCACCAATGAAAGCCACTCTATCAAAG 606

     135 .....................................CysThrAsnTyr 138
                                             |||      :::
     607 CCATCCTGAAAAATATCTCAGTCCTGGCTTTCTCTGTCTGCTTCATCTTC 656
```

```
     139 ThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSe 155
         |||      ::::::   |||||||||   ::::::      |||:::||
     657 ACTATCACCATTGGGATGTTTCCAGCCGTGACTGTTGAGGTCAAGTCCAG 706

     155 r...SerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheL 171
         |   :::|||:::::|||                        :::|||
     707 CATCGCAGGCAGCAGCACCTGG.............GAACGTTACTTCA 741

     171 ysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVal 187
           |||:::||||||                              :::
     742 TTCCTGTGTCCTGT...........................TTCTTG 761

     188 AlaLeuSerLeuPhe.LeuGlyTrpLeuGlyAlaAspArgPheTyr.... 202
         :::::::::::::|||    ||||||   ||||||::::::::::|||
     762 ACTTTCAATATCTTTGACTGGTTGGGCCGGAGCCTCACAGCTGTATTCAT 811

     203 .....LeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGly 217
              ||||||      |||   |||         :::|||      |||
     812 GTGGCCTGGGAAGGACAGCCGCTGGCTGCCAAGCCTGGTGCTGGCCCGGC 861

     218 PheCys 219
         :::|||
     862 TGGTGT 867

seq_name: N_Geneseq_36:V49583

seq_documentation_block:
ID   V49583 standard; cDNA to mRNA; 1119 BP.
AC   V49583;
DT   21-OCT-1998  (first entry)
DE   Human fibrosarcoma cell line HT-1080 clone HP10196 cDNA #2.
KW   Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW   differentiation; immune system; stimulator; suppressor; regulator;
KW   heamatopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW   haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
OS   Homo sapiens.
PN   WO9821328-A2.
PD   22-MAY-1998.
PF   07-NOV-1997; J04056.
PR   13-NOV-1996; JP-301429.
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PI   Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR   WPI; 98-297932/26.
DR   P-PSDB; W64550.
PT   Human protein having transmembrane domain - useful for, e.g.
PT   research and nutrition
PS   Claim 3; Page 128; 205pp; English.
CC   V49550-V49599 are cDNA sequences which encode human proteins containing a
CC   transmembrane domain. These proteins can be used for, e.g. research and
CC   nutrition, and may have cytokine and cell proliferation/differentiation,
CC   immune stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC   activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC   receptor/ligand, anti-inflammatory or tumour inhibition activity.
SQ   Sequence   1119 BP;    203 A;    334 C;    283 G;    299 T;


alignment_scores:
           Quality:  108.00             Length:    236
             Ratio:  0.947                Gaps:     14
 Percent Similarity:  48.305   Percent Identity: 23.305

alignment_block:
 US-09-060-609-2 x V49583   ..

 Align seg 1/1  to: V49583  from: 1  to: 1119

      30 MetArgGlyProPheLysAsnLeuAla...LeuLeuPro......PheSe 43
         :::::::|||   :::   :::|||||||   |||||||||      :::::
     220 CTGCAGGGCAGCCTGTTTGGTCTGGCTGGCCTTCTGCCTGCCAGCTACAC 269

      43 rLeuProLeuLeuGlyGlyGlyGlySerGlySerGlyGluLysValSerV 60
         :   |||::::::::::|||   |||      :::|||      :::|||||

     270 GGCCCCCATCATGAGTGGCCAGGGCCTA...GCAGGCTTCTTTGCCTCCG 316

      60 alSerLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAla 76
         ||:::   |||   ::::||      ||||||   :::      |||:::
     317 TGGCC...ATGATCTGCGCTATTGCCAGTGGCTCGGAGCTATCAGAAAGT 363

      77 ..............ValThrAlaArgLeuValGlyValLeuTrpPheVa 88
                        :::|||||||      |||   ::::||   :::::
     364 GCCTTCGGCTACTTTATCACAGCCTGTGCTGTTATCATTTTGACCATCAT 413

      88 lSerValThrThrGlyPro.....................TrpGlyAlaV 98
         :              |||                     :::      :
     414 CTGTTACCTGGGCCTGCCCCGCCTGGAATTCTACCGCTACTACCAGCAGC 463

      98 alAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLys 114
         ::   :   ::::::   ||||||::::::   |||   :::   ::::::
     464 TCAAGCTTGAAGGACCCGGGGAGCAGGAGACCAAGTTGGACCTCATTAGC 513

     115 ValGlyGlnTyrIleCysLysAspProLys..................... 124
            |||:::             :::|||:::
     514 AAAGGAGAG.............GAGCCAAGAGCAGGCAAAGAGGAATCTGG 551

     125 .IleAsnAspAlaThrGlnGluProValAsn..................... 134
          ::::::   :::::::::::::|||:::|||
     552 AGTTTCAGTCTCCAACTCTCAGCCCACCAATGAAAGCCACTCTATCAAAG 601

     135 .....................................CysThrAsnTyr 138
                                                |||      :::
     602 CCATCCTGAAAAATATCTCAGTCCTGGCTTTCTCTGTCTGCTTCATCTTC 651

     139 ThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSe 155
         |||      ::::::   |||||||||   ::::::      |||:::||
     652 ACTATCACCATTGGGATGTTTCCAGCCGTGACTGTTGAGGTCAAGTCCAG 701

     155 r...SerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheL 171
         |   :::|||:::::|||                        :::|||
     702 CATCGCAGGCAGCAGCACCTGG.............GAACGTTACTTCA 736

     171 ysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVal 187
           |||:::||||||                              :::
     737 TTCCTGTGTCCTGT...........................TTCTTG 756

     188 AlaLeuSerLeuPhe.LeuGlyTrpLeuGlyAlaAspArgPheTyr.... 202
         :::::::::::::|||    ||||||   ||||||::::::::::|||
     757 ACTTTCAATATCTTTGACTGGTTGGGCCGGAGCCTCACAGCTGTATTCAT 806

     203 .....LeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGly 217
              ||||||      |||   |||         :::|||      |||
     807 GTGGCCTGGGAAGGACAGCCGCTGGCTGCCAAGCCTGGTGCTGGCCCGGC 856

     218 PheCys 219
         :::|||
     857 TGGTGT 862

seq_name: N_Geneseq_36:X20370

seq_documentation_block:
ID   X20370 standard; DNA; 715 BP.
AC   X20370;
DT   04-MAY-1999  (first entry)
DE   Borrelia burgdorferi polynucleotide sequence #123.
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
OS   Borrelia burgdorferi.
PN   WO9858943-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; U12764.
PR   03-SEP-1997; US-057483.
PR   20-JUN-1997; US-050359.
PR   22-JUL-1997; US-053344.
PR   22-JUL-1997; US-053377.
PA   (HUMA-) HUMAN GENOME SCI INC.
```

```
PA   (MEDI-) MEDIMMUNE INC.
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI   White OR;
DR   WPI; 99-081217/07.
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
PS   Claim 1; Page 1095-1096; 1128pp; English.
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
SQ   Sequence   715 BP;    268 A;    64 C;    100 G;    281 T;

alignment_scores:
            Quality:  105.00              Length:     63
              Ratio:  2.442                Gaps:      1
 Percent Similarity:  68.254    Percent Identity: 39.683

alignment_block:
 US-09-060-609-2 x X20370   ..

 Align seg 1/1  to: X20370  from: 1  to: 715

     173 IleSerCys............ArgAsnValAsnGlyTyrSerTyrLysVa 185
         |||||||||            ::::::::::::::|||:::      ::
     258 ATTTCTTGCGGAGTCAGAAATAAACAAACCGAAAACTACAATAAACTTAT 307

     185 lAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheT 202
         ::::      |||   |||:::::|||:::||||||:::   |||||||
     308 AGTATTTTTACTATGCTTACTTTTTGGTTATTTAGGAGTTCACAGATTTT 357

     202 yrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPhe 218
         ||:::|||      :::   |||||||||   :::   |||   ||||||
     358 ATGTAGGTAAAATAGGAACTGGTCTATTATACCTATTTACATTTGGATTT 407

     219 CysGlyIleGlySerLeuIleAspPheIleLeuIleSer 231
             :::|||   ||||||||||:::|||   |||:::
     408 TTATATGTTGGAGTTTTAATCGATCTTATTAGAATAACA 446

  name: N_Geneseq_36:V84557

seq_documentation_block:
ID   V84557 standard; DNA; 2127 BP.
AC   V84557;
DT   01-MAR-1999  (first entry)
DE   Human secreted protein gene 147 clone HUSIT49.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9854963-A2.
PD   10-DEC-1998.
PF   04-JUN-1998; U11422.
PR   18-DEC-1997; US-070923.
PR   06-JUN-1997; US-048877.
PR   06-JUN-1997; US-048881.
PR   06-JUN-1997; US-048884.
PR   06-JUN-1997; US-048893.
PR   06-JUN-1997; US-048896.
PR   06-JUN-1997; US-048899.
PR   06-JUN-1997; US-048915.
PR   06-JUN-1997; US-048949.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048972.
PR   06-JUN-1997; US-049020.
PR   06-JUN-1997; US-049375.
PR   05-SEP-1997; US-057628.
PR   05-SEP-1997; US-057635.
PR   05-SEP-1997; US-057644.
PR   05-SEP-1997; US-057647.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057661.
PR   05-SEP-1997; US-057667.
PR   05-SEP-1997; US-057761.
PR   05-SEP-1997; US-057764.
PR   05-SEP-1997; US-057770.
PR   05-SEP-1997; US-057775.
PR   05-SEP-1997; US-057778.
PR   06-JUN-1997; US-048875.
PR   06-JUN-1997; US-048878.
PR   06-JUN-1997; US-048882.
PR   06-JUN-1997; US-048885.
PR   06-JUN-1997; US-048894.
PR   06-JUN-1997; US-048897.
PR   06-JUN-1997; US-048900.
PR   06-JUN-1997; US-048916.
PR   06-JUN-1997; US-048962.
PR   06-JUN-1997; US-048970.
PR   06-JUN-1997; US-048974.
PR   06-JUN-1997; US-049373.
PR   05-SEP-1997; US-057584.
PR   05-SEP-1997; US-057629.
PR   05-SEP-1997; US-057642.
PR   05-SEP-1997; US-057645.
PR   05-SEP-1997; US-057648.
PR   05-SEP-1997; US-057651.
PR   05-SEP-1997; US-057662.
PR   05-SEP-1997; US-057668.
PR   05-SEP-1997; US-057762.
PR   05-SEP-1997; US-057765.
PR   05-SEP-1997; US-057771.
PR   05-SEP-1997; US-057776.
PR   06-JUN-1997; US-048876.
PR   06-JUN-1997; US-048880.
PR   06-JUN-1997; US-048883.
PR   06-JUN-1997; US-048892.
PR   06-JUN-1997; US-048895.
PR   06-JUN-1997; US-048898.
PR   06-JUN-1997; US-048901.
PR   06-JUN-1997; US-048917.
PR   06-JUN-1997; US-048963.
PR   06-JUN-1997; US-048971.
PR   06-JUN-1997; US-049019.
PR   06-JUN-1997; US-049374.
PR   05-SEP-1997; US-057627.
PR   05-SEP-1997; US-057634.
PR   05-SEP-1997; US-057643.
PR   05-SEP-1997; US-057646.
PR   05-SEP-1997; US-057649.
PR   05-SEP-1997; US-057654.
PR   05-SEP-1997; US-057666.
PR   05-SEP-1997; US-057760.
PR   05-SEP-1997; US-057763.
PR   05-SEP-1997; US-057769.
PR   05-SEP-1997; US-057774.
PR   05-SEP-1997; US-057777.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI   Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI   Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI   Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI   Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR   WPI; 99-059865/05.
DR   P-PSDB; W88680, W89005, W89006, W89007, W89008, W89009.
```

```
PT   New isolated human genes and the secreted polypeptides they encode -
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 4; Page 408-409; 772pp; English.
CC   The invention relates to nucleic acid sequences (V84411 to V84633)
CC   encoding human secreted proteins (W88534 to W88756). The secreted protein
CC   gene sequences are deposited with the ATCC under deposit numbers ATCC
CC   97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC   209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC   cells comprising recombinant vectors containing the nucleic acid
CC   sequences are used for the recombinant production of the secreted
CC   proteins. The polynucleotide and amino acid sequences are useful for are
CC   useful for preventing, treating or ameliorating medical conditions e.g.
CC   by protein or gene therapy. Pathological conditions can be also diagnosed
CC   by determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the polynucleotides, based on .
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, neurodegenerative
CC   disorders, developmental abnormalities and foetal deficiencies, blood
CC   disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC   diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC   ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC   restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC   such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC   or thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
CC   The present sequence represents a gene encoding a human secreted protein
CC   (see descriptor line for gene number and clone identification).
SQ   Sequence   2127 BP;    393 A;    618 C;    552 G;    560 T;


alignment_scores:
            Quality:   96.00              Length:    231
              Ratio:  0.906                Gaps:     13
 Percent Similarity:  45.887   Percent Identity: 22.944

alignment_block:
 US-09-060-609-2 x V84557   ..

 Align seg 1/1  to: V84557  from: 1  to: 2127

      37 LeuAlaLeuLeuProPheSerLeu..............ProLeuLeuGl 48
         |||:::   |||   |||   |||               |||::::::::
     603 TTGGTCTGGCTGGCCTTCTGCCTGCCAGCTRACACGGCCCCCATCATGAG 652

      48 yGlyGlyGlySerGlySerGlyGluLysValSerValSerLysMetAlaA 65
         :|||   |||      :::|||      :::|||||||:::   |||   :
     653 TGGCCAGGGCCTA...GCAGGCTTCTTTGCCTCCGTGGCC...ATGATCT 696

      65 laAlaTrpProSerGlyProSerAlaProGluAla.............. 76
         ::|||      ||||||   :::      |||:::
     697 GCGCTATTGCCAGTGGCTCGGAGCTATCAGAAAGTGCCTTCGGCTACTTT 746

      77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGl 93
         :::||||||      |||   :::|||   ::::::
     747 ATCACAGCCTGTGCTGTKATCATTTTGACCATCATCTGTTACCTGGGCCT 796

      93 yPro.....................TrpGlyAlaValAlaThrSerAlaG 103
          |||                    :::      :::      ::::::
     797 GCCCCGCCTGGAATTCTACCGCTACTACCAGCAGCTCAAGCTTGAAGGAC 846

     103 lyGlyGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIle 119
           ||||||::::::   |||   :::   ::::::   |||:::
     847 CCGGGGAGCAGGAGACCAAGTTGGACCTCATTAGCAAAGGAGAG...... 890

     120 CysLysAspProLys....................IleAsnAspAlaTh 129
               :::|||:::                  ::::::   :::::
     891 ......GAGCCAAGAGCAGGCAAAGAGGAATCTGGAGTTTCAGTCTCCAA 934

     129 rGlnGluProValAsn................................ 134
         :::::::|||:::|||
     935 CTCTCAGCCCACCAATGAAAGCCACTCTATCAAAGCCATCCTGAAAAATA 984

     135 .....................CysThrAsnTyrThrAlaHisValSer 143
                              |||      :::|||      ::::::
     985 TCTCAGTCCTGGCTTTCTCTGTCTGCTTCATCTTCACTATCACCATTGGG 1034

     144 CysPheProAlaProAsnIleThrCysLysAspSer...SerGlyAsnGl 159
           |||||||||   ::::::      |||:::|||   :::|||::::::
    1035 ATGTTTCCAGCCGTGACTGTTGAGGTCAAGTCCAGCATCGCAGGCAGCAG 1084

     159 uThrHisPheThrGlyAsnGluValGlyPhePheLysProIleSerCysA 176
         :|||                        :::|||   |||:::||||||
    1085 CACCTGG.............GAACGTTACTTCATTCCTGTGTCCTGT. 1118

     176 rgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPhe 192
                                      :::::::::::::::|||
    1119 ........................TTCTTGACTTTCAATATCTTT 1139

     193 .LeuGlyTrpLeuGlyAlaAspArgPheTyr.........LeuGlyTyrP 206
          ||||||   ||||||:::::::::|||          ||||||
    1140 GACTGGTTGGGCCGGAGCCTCACAGCTGTATTCATGTGGCCTGGGAAGGA 1189

     206 roAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCys 219
           |||   |||           |||      |||:::|||
    1190 CAGCCGCTGGCTGCCAAGCTGGNTGCTGGCCCGGCTGGTGT 1230

seq_name: N_Geneseq_36:T39751

seq_documentation_block:
ID   T39751 standard; cDNA; 1539 BP.
AC   T39751;
DT   01-JAN-1997  (first entry)
DE   Human mucosal addressin cell adhesion molecule-1 cDNA clone 20.
KW   Mucosal addressin cell adhesion molecule-1; MAdCAM; antibody;
KW   leukocyte infiltration; ulcerative colitis; Crohn's disease; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1. .1149
FT                   /*tag=  a
FT   signal_peptide  1. .54
FT                   /*tag=  b
FT   mat_peptide     55. .1146
FT                   /*tag=  c
PN   WO9624673-A1.
PD   15-AUG-1996.
PF   12-FEB-1996; U02153.
PR   10-FEB-1995; US-386857.
PR   01-SEP-1995; US-523004.
PA   (LEUK-) LEUKOSITE INC.
PI   Briskin MJ,  Newman W,  Picarella D,  Ringler DJ;
DR   WPI; 96-384445/38.
DR   P-PSDB; W05322.
PT   Nucleic acid encoding primate mucosal addressin cell adhesion
PT   molecule-1 - and antibodies against this protein, useful to treat
PT   diseases associated with leukocyte infiltration e.g. ulcerative
PT   colitis
PS   Claim 3; Page 108-110; 156pp; English.
CC   cDNA clone 20 (T39751) codes for human mucosal addressin cell
CC   adhesion molecule-1 (MAdCAM-1) (W05322), whose cellular adhesion
CC   function is alpha4beta7 integrin-dependent and pref. is selective
CC   for alpha4beta7.  It was isolated from a cDNA library prepd. from
CC   mesenteric lymph nodes (MLN) of a patient with Crohn's disease,
CC   using a macaque MAdCAM-1 cDNA clone (T39752) as probe.  Another
CC   clone (T39750), encoding a different human MAdCAM isoform (W05321),
CC   was isolated from a histologically normal MLN library.  MAdCAM
CC   nucleic acids can be used as probes to detect polymorphic or
CC   allelic variants of MAdCAM, or used to produce recombinant MAdCAMs
CC   in host cells.
SQ   Sequence   1539 BP;    268 A;    570 C;    442 G;    259 T;


alignment_scores:
            Quality:   91.50              Length:    166
              Ratio:  1.102                Gaps:      9
```

```
 Percent Similarity:  50.000   Percent Identity: 25.904

alignment_block:
 US-09-060-609-2 x T39751/rev   ..

 Align seg 1/1  to reverse of: T39751  from: 1  to: 1539

       4 LeuLysGlySerProAsnValIleProArgAlaHis...GlyGlnLysAs 19
         |||   |||   |||         |||      |||   |||||||||
    1128 CTGGCCGGTCCCCCTTAACCCAGCCCAGGCCGACACCTGGGGCAGAAGCC 1079

      19 nThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProPheLysA 36
             |||         :::|||:::
    1078 TCAGAGAAGCTGGTGGGTGGGTGTCGTC..................... 1051

      36 snLeuAlaLeuLeuProPheSerLeuPro...LeuLeuGlyGlyGlyGly 51
           |||:::   :::|||   ::::::|||   ::::::||||||   |||
    1050 ..CTCAGCCAGGTGCCGGCAGCGTTTCCAGAGGTGATAGGTGGGCAAGGC 1003

      52 Ser.GlySerGlyGluLysValSer......ValSerLysMetAlaAlaA 66
         ::: :::|||:::   :::::::::      |||   :::   |||:::: 
    1002 CAGGAGCAGCAGTCCCAGCACCGCACTGCTGGTCCACAGAGCCGCGGGCA 953

      66 laTrp...ProSerGlyProSerAlaProGluAlaValThrAlaArgLeu 81
         ::|||   |||:::|||   :::   |||   ::::::|||:::
     952 GCTGGTCACCCGCAGGTTTGGACGAGCCTGTTGGGATCACTTCTCCCTGC 903

      82 ValGlyValLeuTrpPheValSer..............Val.ThrThrG 93
         ||||||      |||   :::|||                ||| :::   |
     902 GTGGGCCCAGCCTGGGAGATCTCAGGGCGGCGAGTCCTGGTGGAGCCTGG 853

      93 lyProTrpGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLeuLys 109
         ||   ||||||:::|||   :::   ||||||||||:::         :::
     852 GCTCCTGGGGGTGTGTGTGGAGCCCTGCTGGGGGGCGGGCTCCGGGGAGG 803

     110 CysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAs 126
                                  |||::::::         ::
     802 TCT.........................TGTCGGGAGGCTCCGGGGAG 780

     126 nAspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValS 143
         :   :::   :::   |||   :::      :::   |||
     779 GTGGTGTCGGGAGGCTCCTGGGAGGTGGTGTCGGGAGACTCCGGGGAGGT 730

     143 erCysPheProAlaProAsnIleThrCysLysAspSerSerGly 157
           |||      :::|||:::      |||:::::::::   |||
     729 GGTGTCGGGAGACTCCGGGGAGGTGGTGTCGGGAGGCTCCGGGG 686

 Name: N_Geneseq_36:X35280

seq_documentation_block:
ID   X35280 standard; DNA; 1539 BP.
AC   X35280;
DT   02-JUL-1999  (first entry)
DE   cDNA encoding human MAdCAM-1.
KW   Mucosal addressing cell adhesion molecule-1; MAdCAM-1; inhibitor;
KW   leukocyte infiltration; disease; inflammatory bowel disease;
KW   insulin-dependent diabetes mellitus; inflammatory bowel disease;
KW   ulcerative colitis; Crohn's disease; ileitis' coeliac disease;
KW   nontropical Sprue; enteropathy; seronegative arthropathy; colitis;
KW   eosinophilic gastroenteritis; pouchitis; proctocolectomy;
KW   ileoanal anastomosis; alpha4-beta-integrin; ss.
OS   Homo sapiens.
PN   WO9725351-A2.
PD   17-JUL-1997.
PF   03-JAN-1997; U00291.
PR   04-JAN-1996; US-582740.
PA   (LEUK-) LEUKOSITE INC.
PI   Schwender CF, Shroff HN;
DR   WPI; 97-384978/35.
DR   P-PSDB; Y02069.
PT   New polypeptide(s) for inhibiting MAdCAM-1-mediated interactions -
PT   useful for treatment of diseases associated with leukocyte
PT   infiltration of tissues, especially inflammatory bowel disease and
PT   insulin-dependent diabetes mellitus
PS   Disclosure; Fig 5; 108pp; English.
CC   The specification describes inhibitors of mucosal addressing cell
CC   adhesion molecule-1 (MAdCAM-1) mediated interactions. These inhibitors
CC   comprise peptides which mimic the conserved amino acid motif LDTSL of
CC   MAdCAM-1, and which have groups bonded to the N and C terminals.
CC   The inhibitors can be used for the treatment of a disease associated
CC   with leukocyte infiltration of tissues expressing the MAdCAM-1,
CC   especially inflammatory bowel disease and insulin-dependent diabetes
CC   mellitus. Inflammatory bowel diseases which can be treated include
CC   ulcerative colitis, Crohn's disease, ileitis, coeliac disease,
CC   nontropical Sprue, enteropathy associated with seronegative
CC   arthropathies, microscopic or collagenous colitis, eosinophilic
CC   gastroenteritis, or pouchitis resulting after proctocolectomy, and
CC   ileoanal anastomosis. The inhibitors can also be used to inhibit the
CC   bonding of a cell expressing a ligand for MAdCAM-1 on the cell surface to
CC   MAdCAM-1. They are also useful in diagnostic and research applications,
CC   e.g. as immunogens (when conjugated to a suitable carrier) to induce the
CC   formation of antibodies which selectively bind MAdCAM-1. These antibodies
CC   can in turn be used to identify cells expressing MAdCAM-1 on their cell
CC   surface or to detect MAdCAM-1 in a sample. The compounds can also be
CC   labelled and used to detect alpha4-beta-integrin and/or to quantitate
CC   expression of this the surface of cells. The present sequence
CC   encodes human MAdCAM-1.
SQ   Sequence   1539 BP;    268 A;    570 C;    442 G;    259 T;


alignment_scores:
           Quality:   91.50            Length:    166
             Ratio:  1.102              Gaps:      9
 Percent Similarity:  50.000   Percent Identity: 25.904

alignment_block:
 US-09-060-609-2 x X35280/rev   ..

 Align seg 1/1  to reverse of: X35280  from: 1  to: 1539

       4 LeuLysGlySerProAsnValIleProArgAlaHis...GlyGlnLysAs 19
         |||   |||   |||         |||      |||   |||||||||
    1128 CTGGCCGGTCCCCCTTAACCCAGCCCAGGCCGACACCTGGGGCAGAAGCC 1079

      19 nThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProPheLysA 36
             |||         :::|||:::
    1078 TCAGAGAAGCTGGTGGGTGGGTGTCGTC..................... 1051

      36 snLeuAlaLeuLeuProPheSerLeuPro...LeuLeuGlyGlyGlyGly 51
           |||:::   :::|||   ::::::|||   ::::::||||||   |||
    1050 ..CTCAGCCAGGTGCCGGCAGCGTTTCCAGAGGTGATAGGTGGGCAAGGC 1003

      52 Ser.GlySerGlyGluLysValSer......ValSerLysMetAlaAlaA 66
         ::: :::|||:::   :::::::::      |||   :::   |||:::: 
    1002 CAGGAGCAGCAGTCCCAGCACCGCACTGCTGGTCCACAGAGCCGCGGGCA 953

      66 laTrp...ProSerGlyProSerAlaProGluAlaValThrAlaArgLeu 81
         ::|||   |||:::|||   :::   |||   ::::::|||:::
     952 GCTGGTCACCCGCAGGTTTGGACGAGCCTGTTGGGATCACTTCTCCCTGC 903

      82 ValGlyValLeuTrpPheValSer..............Val.ThrThrG 93
         ||||||      |||   :::|||                ||| :::   |
     902 GTGGGCCCAGCCTGGGAGATCTCAGGGCGGCGAGTCCTGGTGGAGCCTGG 853

      93 lyProTrpGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLeuLys 109
         ||   ||||||:::|||   :::   ||||||||||:::         :::
     852 GCTCCTGGGGGTGTGTGTGGAGCCCTGCTGGGGGGCGGGCTCCGGGGAGG 803

     110 CysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAs 126
                                  |||::::::         ::
     802 TCT.........................TGTCGGGAGGCTCCGGGGAG 780

     126 nAspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValS 143
         :   :::   :::   |||   :::      :::   |||
     779 GTGGTGTCGGGAGGCTCCTGGGAGGTGGTGTCGGGAGACTCCGGGGAGGT 730
```

```
     143 erCysPheProAlaProAsnIleThrCysLysAspSerSerGly 157
         |||      :::|||:::      |||:::::::::   |||
     729 GGTGTCGGGAGACTCCGGGGAGGTGGTGTCGGGAGGCTCCGGGG 686

seq_name: N_Geneseq_36:V22918

seq_documentation_block:
ID   V22918 standard; cDNA; 2397 BP.
AC   V22918;
DT   29-JUL-1998  (first entry)
DE   Human GlyT-2 transporter G1387A mutant gene sequence.
KW   Human; GlyT-2 transporter; glycine transporter; neurotransmitter;
KW   activity; N-methyl-D-aspartate receptor; NMDA;
KW   strychnine-sensitive glycine receptor; treatment; diagnosis;
KW   nervous system disorder; myoclonus; muscle spasm; hyperactivity;
KW   epilepsy; spasticity; head trauma; neuronal cell death;
KW   multiple sclerosis; spinal cord injury; Huntington's disease;
KW   amyotrophic lateral sclerosis; ds.
OS   Synthetic.
OS   Homo sapiens.
FT   Key             Location/Qualifiers
FT   mutation        1387
FT                   /*tag=  a
FT                   /note= "G1387 replaced with A"
PN   WO9807854-A1.
PD   26-FEB-1998.
PF   20-AUG-1997; U14637.
PR   20-AUG-1996; US-700013.
PA   (ALLX ) ALLELIX NEUROSCIENCE INC.
PI   Albert VR, Borden LA, Kowalski LR, McKelvy JF;
DR   WPI; 98-169158/15.
PT   Human glycine transporter - useful for treating nervous system
PT   disorders, e.g. pain, myoclonus, etc.
PS   Claim 6; Page -; 90pp; English.
CC   The present sequence represents a human GlyT-2 transporter mutant
CC   gene sequence. GlyT-2 is a glycine transporter found predominantly in
CC   the brain stem and the spinal cord. GlyT-2 regulates the levels of
CC   glycine, a major neurotransmitter, that preferentially influences the
CC   activity of N-methyl-D-aspartate (NMDA) receptors and
CC   strychnine-sensitive glycine receptors. The GlyT-2 protein can be
CC   used to treat or to diagnose a nervous system disorder or condition,
CC   e.g. pain, myoclonus, muscle spasm or hyperactivity, epilepsy or
CC   spasticity associated with stroke, head trauma, neuronal cell death,
CC   multiple sclerosis, spinal cord injury, dystonia, Huntington's disease
CC   or amyotrophic lateral sclerosis.
CC   note: the present sequence does not appear in the specification; it was
CC   created using information provided.
SQ   Sequence   2397 BP;    509 A;    665 C;    628 G;    595 T;

alignment_scores:
            Quality:   88.00              Length:    302
              Ratio:  0.624               Gaps:       17
 Percent Similarity:  46.689   Percent Identity: 22.517

alignment_block:
 US-09-060-609-2 x V22918   ..

 Align seg 1/1  to: V22918  from: 1  to: 2397

      25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
         |||||||||:::|||::::::|||      ::::::   |||
     591 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 640

      39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly.............. 51
          |||      |||||||||   |||         |||
     641 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 690

      52 .................SerGlySerGly.Glu................. 56
                         :::||||||||||
     691 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 740

      57 .LysValSerValSerLysMetAlaAlaAlaTrpPro.............. 68

         :::||||||:::::::::::|||:::      |||
     741 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 790

      69 ................SerGlyProSerAlaProGluAlaValThrAla 79
                         :::|||   :::          :::   :::
     791 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCATCTCT 840

      80 ArgLeuValGlyValLeuTrp............................ 86
         |||:::::::::   :::
     841 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 890

      87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                     ||||||||||||        |||||||||:::
     891 CCTGTTTGCCTCCTTTGTGTCTGTACTA......CCCTGGGGCTCCTGCA 934

      99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp......... 112
         :::      :::      |||         |||:::|||
     935 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 975

     113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
         |||   :::   :::   :::   :::|||   |||||||||
     976 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA........ 1017

     128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                        :::      |||   |||      ::::::::::
    1018 ................CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1051

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
         ::|||      :::::::::         |||::::::|||:::|||
    1052 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1101

     162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
            :::|||:::|||      :::||||||   :::      :::
    1102 GTCAGTGGAAGTGAA...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1148

     178 lAsnGlyTyrSerTyr............LysValAlaValAlaLeuSerL 191
            |||   :::|||           :::      :::||||||   |
    1149 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1198

     191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
         ||||||||:::|||:::                  ::   |||   :::
    1199 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1230

     208 Leu.....................GlyLeuLeuLysPheCysThrValGl 217
         |||                    |||   :::   :::   |||:::
    1231 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1280

     217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
          |||      :::      |||:::   :::::::::   ::::::   :
    1281 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1327

     234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
         ::   |||   :::   |||      |||:::|||
    1328 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1377

     251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
         :::|||||||:::   ::::::   :::::::::         ||||||::
    1378 AAACTCACGAATGCCACGGTGTGGAAAGATGCT......GCCACTCAGAT 1421

     267 uTyr 268
         ::::
    1422 TTTC 1425

seq_name: N_Geneseq_36:V22914

seq_documentation_block:
ID   V22914 standard; cDNA; 2397 BP.
AC   V22914;
DT   29-JUL-1998  (first entry)
DE   Human GlyT-2 transporter T836A mutant gene sequence.
KW   Human; GlyT-2 transporter; glycine transporter; neurotransmitter;
KW   activity; N-methyl-D-aspartate receptor; NMDA;
KW   strychnine-sensitive glycine receptor; treatment; diagnosis;
```

```
KW   nervous system disorder; myoclonus; muscle spasm; hyperactivity;
KW   epilepsy; spasticity; head trauma; neuronal cell death;
KW   multiple sclerosis; spinal cord injury; Huntington's disease;
KW   amyotrophic lateral sclerosis; ds.
OS   Synthetic.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   mutation        836
FT                   /*tag=  a
FT                   /note= "T836 replaced with A"
PN   WO9807854-A1.
PD   26-FEB-1998.
PF   20-AUG-1997; U14637.
PR   20-AUG-1996; US-700013.
PA   (ALLX ) ALLELIX NEUROSCIENCE INC.
PI   Albert VR, Borden LA, Kowalski LR, McKelvy JF;
DR   WPI; 98-169158/15.
PT   Human glycine transporter - useful for treating nervous system
     disorders, e.g. pain, myoclonus, etc.
     Claim 6; Page -; 90pp; English.
CC   The present sequence represents a human GlyT-2 transporter mutant
CC   gene sequence. GlyT-2 is a glycine transporter found predominantly in
CC   the brain stem and the spinal cord. GlyT-2 regulates the levels of
CC   glycine, a major neurotransmitter, that preferentially influences the
CC   activity of N-methyl-D-aspartate (NMDA) receptors and
CC   strychnine-sensitive glycine receptors. The GlyT-2 protein can be
CC   used to treat or to diagnose a nervous system disorder or condition,
CC   e.g. pain, myoclonus, muscle spasm or hyperactivity, epilepsy or
CC   spasticity associated with stroke, head trauma, neuronal cell death,
CC   multiple sclerosis, spinal cord injury, dystonia, Huntington's disease
CC   or amyotrophic lateral sclerosis.
CC   note: the present sequence does not appear in the specification; it was
CC   created using information provided.
SQ   Sequence   2397 BP;    509 A;    665 C;    629 G;    594 T;


alignment_scores:
           Quality:   87.00              Length:    302
             Ratio:  0.617                Gaps:     17
 Percent Similarity:  46.689   Percent Identity: 22.517

alignment_block:
 US-09-060-609-2 x V22914   ..

 Align seg 1/1  to: V22914  from: 1  to: 2397

    25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
       ||||||||||:::|||:::::::|||      ::::::   |||
   591 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 640

    39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly.............. 51
        |||      |||||||||   |||       |||
   641 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 690

    52 .................SerGlySerGly.Glu................. 56
                        :::|||||||||
   691 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 740

    57 .LysValSerValSerLysMetAlaAlaAlaTrpPro.............. 68
       :::||||||:::::::::::::|||:::      |||
   741 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 790

    69 .................SerGlyProSerAlaProGluAlaValThrAla 79
                        :::|||   :::          :::::::::
   791 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCAACTCT 840

    80 ArgLeuValGlyValLeuTrp............................. 86
        |||:::::::::   :::
   841 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 890

    87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                    ||||||||||||         |||||||||:::
   891 CCTGTTTGCCTCCTTTGTGTCTGTACTA......CCCTGGGGCTCCTGCA 934

    99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp......... 112
         :::      :::      |||         |||:::|||
   935 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 975

   113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
       |||   :::   :::   :::   :::|||   |||||||||
   976 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA........ 1017

   128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                     :::      |||  |||      ::::::::::
  1018 ...............CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1051

   145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
       ::|||      :::::::::         |||:::::|||:::|||
  1052 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1101

   162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
          :::|||:::|||      :::||||||   :::      :::
  1102 GTCAGTGGAAGTGAA...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1148

   178 lAsnGlyTyrSerTyr............LysValAlaValAlaLeuSerL 191
         |||   :::|||              :::      :::||||||   |
  1149 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1198

   191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
       ||||||||:::|||:::                  :::   |||   :::
  1199 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1230

   208 Leu.....................GlyLeuLeuLysPheCysThrValGl 217
       |||                    |||   :::   :::   |||:::
  1231 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1280

   217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
        |||      :::      |||:::   :::::::::   ::::::   :
  1281 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1327

   234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
       ::   |||   :::  |||       |||:::|||
  1328 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1377

   251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
       :::||||||      ::::::   :::::::::      ||||||::
  1378 AAACTCACGGATGCCACGGTGTGGAAAGATGCT......GCCACTCAGAT 1421

   267 uTyr 268
       ::::
  1422 TTTC 1425

seq_name: N_Geneseq_36:V22906

seq_documentation_block:
ID   V22906 standard; cDNA; 2397 BP.
AC   V22906;
DT   29-JUL-1998  (first entry)
DE   Human GlyT-2 variant gene sequence.
KW   Human; GlyT-2 transporter; glycine transporter; neurotransmitter;
KW   activity; N-methyl-D-aspartate receptor; NMDA;
KW   strychnine-sensitive glycine receptor; treatment; diagnosis;
KW   nervous system disorder; myoclonus; muscle spasm; hyperactivity;
KW   epilepsy; spasticity; head trauma; neuronal cell death;
KW   multiple sclerosis; spinal cord injury; Huntington's disease;
KW   amyotrophic lateral sclerosis; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .2394
FT                   /*tag=  a
FT                   /product= GlyT-2 transporter
PN   WO9807854-A1.
PD   26-FEB-1998.
PF   20-AUG-1997; U14637.
PR   20-AUG-1996; US-700013.
PA   (ALLX ) ALLELIX NEUROSCIENCE INC.
```

```
PI   Albert VR, Borden LA, Kowalski LR, McKelvy JF;
DR   WPI; 98-169158/15.
DR   P-PSDB; W56369.
PT   Human glycine transporter - useful for treating nervous system
PT   disorders, e.g. pain, myoclonus, etc.
PS   Example 1A; Pages 47-48; 90pp; English.
CC   The present sequence encodes a variant of the human GlyT-2 transporter.
CC   The cDNA encoding GlyT-2 was generated by reverse transcription
CC   PCR (RT-PCR) from human spinal cord or cerebellum polyA mRNA. GlyT-2 is
CC   a glycine transporter found predominantly in the brain stem and the
CC   spinal cord. GlyT-2 regulates the levels of glycine, a major
CC   neurotransmitter, that preferentially influences the activity of
CC   N-methyl-D-aspartate (NMDA) receptors and strychnine-sensitive glycine
CC   receptors. The GlyT-2 protein can be used to treat or to diagnose a
CC   nervous system disorder or condition, e.g. pain, myoclonus, muscle spasm
CC   or hyperactivity, epilepsy or spasticity associated with stroke, head
CC   trauma, neuronal cell death, multiple sclerosis, spinal cord injury,
CC   dystonia, Huntington's disease or amyotrophic lateral sclerosis.
SQ   Sequence   2397 BP;    509 A;    662 C;    629 G;    597 T;

gnment_scores:
          Quality:   87.00            Length:    302
            Ratio:  0.617              Gaps:     17
 Percent Similarity:  46.689   Percent Identity: 22.517

alignment_block:
 US-09-060-609-2 x V22906   ..

 Align seg 1/1  to: V22906  from: 1  to: 2397

      25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
         ||||||||||:::|||:::::::|||      ::::::   |||
     591 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 640

      39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly.............. 51
          |||      |||||||||   |||         |||
     641 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 690

      52 .................SerGlySerGly.Glu................... 56
                         :::||||||||||
     691 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 740

      57 .LysValSerValSerLysMetAlaAlaAlaTrpPro.............. 68
         :::|||||:::::::::::::|||:::      |||
     741 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 790

      69 .................SerGlyProSerAlaProGluAlaValThrAla 79
                         :::|||   :::          :::::::::
     791 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCAACTCT 840

      80 ArgLeuValGlyValLeuTrp............................. 86
          |||:::::::::   :::
     841 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 890

      87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                     ||||||||||||         |||||||||:::
     891 CCTGTTTGCCTCCTTTGTGTCTGTACTA......CCCTGGGGCTCCTGCA 934

      99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp......... 112
         :::      :::      |||         |||:::|||
     935 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 975

     113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
         |||   :::   :::   :::   :::|||   |||||||||
     976 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA........ 1017

     128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                    :::      |||   |||     ::::::::::
    1018 ...............CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1051

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
         ::|||      :::::::::         |||:::::::|||:::|||

    1052 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1101

     162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
            :::|||:::|||      :::||||||   :::      :::
    1102 GTCAGTGGAAGTGAG...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1148

     178 lAsnGlyTyrSerTyr............LysValAlaValAlaLeuSerL 191
          |||   :::|||           :::      :::||||||   |
    1149 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1198

     191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
         ||||||||:::|||:::                  :::   |||   :::
    1199 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1230

     208 Leu.....................GlyLeuLeuLysPheCysThrValGl 217
         |||                    |||   :::   :::   |||:::
    1231 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1280

     217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
          |||      :::      |||:::   :::::::::   ::::::   :
    1281 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1327

     234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
         ::   |||   :::   |||      |||:::|||
    1328 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1377

     251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
         :::|||||      ::::::   :::::::::         ||||||::
    1378 AAACTCACGGATGCCACGGTGTCGAAAGATGCT......GCCACTCAGAT 1421

     267 uTyr 268
         ::::
    1422 TTTC 1425

seq_name: N_Geneseq_36:V22900

seq_documentation_block:
ID   V22900 standard; cDNA; 1597 BP.
AC   V22900;
DT   29-JUL-1998  (first entry)
DE   Fragment of the human GlyT-2 gene sequence in clone phG2-2.
KW   Human; GlyT-2 transporter; glycine transporter; neurotransmitter;
KW   activity; N-methyl-D-aspartate receptor; NMDA;
KW   strychnine-sensitive glycine receptor; treatment; diagnosis;
KW   nervous system disorder; myoclonus; muscle spasm; hyperactivity;
KW   epilepsy; spasticity; head trauma; neuronal cell death;
KW   multiple sclerosis; spinal cord injury; Huntington's disease;
KW   amyotrophic lateral sclerosis; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             2. .535
FT                   /*tag= a
FT   CDS             536. .1597
FT                   /*tag= b
FT                   /note= "no stop codon given"
PN   WO9807854-A1.
PD   26-FEB-1998.
PF   20-AUG-1997; U14637.
PR   20-AUG-1996; US-700013.
PA   (ALLX ) ALLELIX NEUROSCIENCE INC.
PI   Albert VR, Borden LA, Kowalski LR, McKelvy JF;
DR   WPI; 98-169158/15.
DR   P-PSDB; W56362, W56363.
PT   Human glycine transporter - useful for treating nervous system
PT   disorders, e.g. pain, myoclonus, etc.
PS   Example 1A; Pages 39-40; 90pp; English.
CC   The present sequence encodes a fragment of the human GlyT-2 transporter.
CC   It corresponds to nucleotides 39 to 1635 of the full length sequence
CC   V22909. The cDNA encoding GlyT-2 was generated by reverse transcription
CC   PCR (RT-PCR) from human spinal cord or cerebellum polyA mRNA, using
CC   PCR primers V22939-40. The sequence differs slightly from the full length
CC   GlyT-2 sequence because samples were collected from several donors.
CC   GlyT-2 is a glycine transporter found predominantly in the brain stem
```

```
CC   and the spinal cord. GlyT-2 regulates the levels of glycine, a major
CC   neurotransmitter, that preferentially influences the activity of
CC   N-methyl-D-aspartate (NMDA) receptors and strychnine-sensitive glycine
CC   receptors. The GlyT-2 protein can be used to treat or to diagnose a
CC   nervous system disorder or condition, e.g. pain, myoclonus, muscle spasm
CC   or hyperactivity, epilepsy or spasticity associated with stroke, head
CC   trauma, neuronal cell death, multiple sclerosis, spinal cord injury,
CC   dystonia, Huntington's disease or amyotrophic lateral sclerosis.
SQ   Sequence   1597 BP;    338 A;    457 C;    447 G;    355 T;


alignment_scores:
            Quality:   87.00              Length:    302
              Ratio:  0.617                Gaps:     17
 Percent Similarity:  46.689   Percent Identity: 22.517

alignment_block:
US-09-060-609-2 x V22900   ..

Align seg 1/1  to: V22900  from: 1  to: 1597

     25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
        |||||||||:::|||:::::::|||      ::::::   |||
    553 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 602

     39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly..... 51
         |||      |||||||||   |||         |||
    603 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 652

     52 .................SerGlySerGly.Glu..... 56
                        :::|||||||||
    653 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 702

     57 .LysValSerValSerLysMetAlaAlaAlaTrpPro..... 68
         :::||||||:::::::::::::|||:::      |||
    703 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 752

     69 .................SerGlyProSerAlaProGluAlaValThrAla 79
                        :::|||   :::           :::::::::
    753 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCAACTCT 802

     80 ArgLeuValGlyValLeuTrp..... 86
         |||::::::::::   :::
    803 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 852

     87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                    ||||||||||||         |||||||||:::
    853 CCTGTTTGCCTCCTTTGTGTCTGTACTA......CCCTGGGGCTCCTGCA 896

     99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp..... 112
          :::      :::      |||         |||:::|||
    897 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 937

    113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
        |||   :::   :::   :::   :::|||   |||||||||
    938 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA........ 979

    128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                      :::      |||   |||      ::::::::::
    980 ................CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1013

    145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
        ::|||      :::::::::         |||::::::|||:::|||
   1014 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1063

    162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
          :::|||:::|||      :::||||||   :::       :::
   1064 GTCAGTGGAAGTGAG...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1110

    178 lAsnGlyTyrSerTyr.....LysValAlaValAlaLeuSerL 191
           |||   :::|||          :::      :::||||||   |
   1111 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1160

    191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
        ||||||||:::|||:::                  :::   |||   :::
   1161 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1192

    208 Leu.....................GlyLeuLeuLysPheCysThrValGl 217
        |||                    |||   :::   :::   |||:::
   1193 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1242

    217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
         |||      :::      |||:::   :::::::::   ::::::   :
   1243 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1289

    234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
        ::   |||   :::   |||      |||:::|||
   1290 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1339

    251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
        :::|||||      ::::::   :::::::::         ||||||::
   1340 AAACTCACGGATGCCACGGTGTGGAAAGATGCT......GCCACTCAGAT 1383

    267 uTyr 268
        ::::
   1384 TTTC 1387

seq_name: N_Geneseq_36:T39750

seq_documentation_block:
ID   T39750 standard; cDNA; 1624 BP.
AC   T39750;
DT   01-JAN-1997  (first entry)
DE   Human mucosal addressin cell adhesion molecule-1 cDNA clone 4.
KW   Mucosal addressin cell adhesion molecule-1; MAdCAM; antibody;
KW   leukocyte infiltration; ulcerative colitis; Crohn's disease; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1. .1221
FT                   /*tag=  a
FT   signal_peptide  1. .54
FT                   /*tag=  b
FT   mat_peptide     55. .1218
FT                   /*tag=  c
PN   WO9624673-A1.
PD   15-AUG-1996.
PF   12-FEB-1996; U02153.
PR   10-FEB-1995; US-386857.
PR   01-SEP-1995; US-523004.
PA   (LEUK-) LEUKOSITE INC.
PI   Briskin MJ,  Newman W,  Picarella D,  Ringler DJ;
DR   WPI; 96-384445/38.
DR   P-PSDB; W05321.
PT   Nucleic acid encoding primate mucosal addressin cell adhesion
PT   molecule-1 - and antibodies against this protein, useful to treat
PT   diseases associated with leukocyte infiltration e.g. ulcerative
PT   colitis
PS   Claim 3; Page 104-106; 156pp; English.
CC   cDNA clone 4 (T39750) codes for human mucosal addressin cell
CC   adhesion molecule-1 (MAdCAM-1) (W05321), whose cellular adhesion
CC   function is alpha4beta7 integrin-dependent and pref. is selective
CC   for alpha4beta7.  It was isolated from a cDNA library prepd. from
CC   normal human mesenteric lymph nodes (MLN) using a macaque MAdCAM-1
CC   cDNA clone (T39752) as probe.  Another clone (T39751), encoding
CC   a different human MAdCAM isoform (W05322), was isolated from the
CC   MLN of a patient with Crohn's disease.  MAdCAM nucleic acids can
CC   be used as probes to detect polymorphic or allelic variants of
CC   MAdCAM, or used to produce recombinant MAdCAMs in host cells.
SQ   Sequence   1624 BP;    295 A;    610 C;    454 G;    265 T;


alignment_scores:
            Quality:   86.50              Length:    182
              Ratio:  0.961                Gaps:     11
 Percent Similarity:  49.451   Percent Identity: 24.725
```

```
alignment_block:
 US-09-060-609-2 x T39750/rev   ..

 Align seg 1/1  to reverse of: T39750  from: 1  to: 1624

      4 LeuLysGlySerProAsnValIleProArgAlaHis...GlyGlnLysAs 19
        |||   |||   |||         |||      |||   |||||||||
   1200 CTGGCCGGTCCCCCTTAACCCAGCCCAGGCCGACACCTGGGGCAGAAGCC 1151

     19 nThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProPheLysA 36
          |||         :::|||:::
   1150 TCAGAGAAGCTGGTGGGTGGGTGTCGTC..................... 1123

     36 snLeuAlaLeuLeuProPheSerLeuPro...LeuLeuGlyGlyGlyGly 51
          |||:::   :::|||   ::::::|||   ::::::   |||   |||
   1122 ..CTCAGCCAGGTGCCGGCAGCGTTTCCAGAGGTGATACGTGGGCAAGGC 1075

     52 Ser.GlySerGlyGluLysValSer......ValSerLysMetAlaAlaA 66
        ::: :::|||:::   :::::::::      |||   :::   |||::::
   1074 CAGGAGCAGCAGTCCCAGCACCGCACTGCTGGTCCACAGAGCCGCGGGCA 1025

     66 laTrp...ProSerGlyProSerAlaProGluAlaValThrAlaArgLeu 81
        ::|||   |||:::|||   :::   |||   ::::::|||:::
   1024 GCTGGTCACCCGCAGGTTTGGACGAGCCTGTTGGGATCACTTCTCCCTGC 975

     82 ValGlyValLeuTrpPheValSer..............Val.ThrThrG 93
        ||||||      |||   :::|||               ||| :::   |
    974 GTGGGCCCAGCCTGGGAGATCTCAGGGCGGCGAGTCCTGGTGGAGCCTGG 925

     93 lyProTrpGlyAlaValAlaThrSerAlaGlyGlyGlu...GluSerLeu 108
        ||   ||||||:::|||   :::   |||||||||:::       :::
    924 GCTCCTGGGGGTGTGTGTGGAGCCCTGCTGGGGGGCGGGCTCCGGGGAGG 875

     109 LysCysGluAspLeuLysValGlyGlnTyrIle..................119
         :::|||:::::::          |||::::::
     874 TCTTGTCGGGAGGCTCCG...GGGAGGTGGTGTCGGGAGGCTCCTGGGAG 828

     120 ............................CysLysAspProLysIleAsnA 127
                                     |||::::::       :::
     827 GTGGTGTCGGGAGGCTCCTGGGAGGTGGTGTCGGGAGGCTCCTGGGAGGT 778

     127 spAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSer 143
           :::   :::   |||   :::      :::   |||
     777 GGTGTCGGGAGACTCCGGGGAGGTGGTGTCGGGAGACTCCGGGGAGGTGG 728

     144 CysPheProAlaProAsnIleThrCysLysAspSerSerGly 157
         |||:::   ||||||:::      |||:::::::::   |||
     727 TGTTGGGAGGCTCCGGGGAGGTGGTGTCGGGAGGCTCCGGGG 686
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 13:23:38 ; Search time 42.25 Seconds
(without alignments)
2220.775 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: IDENTITY_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
	1	36.4	4.5	1462	2	US-08-552-142A-16	Sequence 16, Appl
	2	36.4	4.5	1951	2	US-08-910-973-16	Sequence 16, Appl
c	3	34.6	4.3	4673	1	US-07-638-431-1	Sequence 1, Appli
c	4	34.6	4.3	4673	5	PCT-US92-00018-1	Sequence 1, Appli
	5	33.4	4.1	1636	6	5447867-2	Patent No. 5447867
c	6	32.8	4.0	988	1	US-08-243-545-5	Sequence 5, Appli
	7	32.8	4.0	1700	2	US-08-539-304A-3	Sequence 3, Appli
c	8	32.8	4.0	988	3	US-08-993-962-5	Sequence 5, Appli
c	9	32.8	4.0	988	5	PCT-US94-05365-5	Sequence 5, Appli
	10	32.4	4.0	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	11	32.2	4.0	3012	3	US-08-475-427-5	Sequence 5, Appli
c	12	32.2	4.0	3012	4	US-07-842-165-5	Sequence 5, Appli
c	13	31.6	3.9	532	1	US-08-558-735-1	Sequence 1, Appli
c	14	31.6	3.9	345	1	US-08-558-735-3	Sequence 3, Appli
c	15	31.6	3.9	355	1	US-08-558-735-16	Sequence 16, Appl
c	16	31.6	3.9	345	1	US-08-558-735-17	Sequence 17, Appl
c	17	31.6	3.9	333	1	US-08-558-735-18	Sequence 18, Appl
c	18	31.6	3.9	330	1	US-08-558-735-19	Sequence 19, Appl
c	19	31.6	3.9	857	5	PCT-US94-05150-19	Sequence 19, Appl
	20	31.4	3.9	8700	3	US-08-392-625-16	Sequence 16, Appl
	21	31.4	3.9	8700	3	US-08-466-961A-16	Sequence 16, Appl
	22	31.4	3.9	8700	4	US-08-645-193B-18	Sequence 18, Appl
c	23	31.4	3.9	8298	5	PCT-US93-03076-1	Sequence 1, Appli
c	24	31.2	3.9	1644	3	US-08-458-555-1	Sequence 1, Appli
	25	31	3.8	2688	4	US-08-909-965C-1	Sequence 1, Appli
c	26	30.6	3.8	440	1	US-08-558-735-4	Sequence 4, Appli
c	27	30.6	3.8	345	1	US-08-558-735-6	Sequence 6, Appli
c	28	30.6	3.8	522	1	US-08-558-735-7	Sequence 7, Appli
c	29	30.6	3.8	342	1	US-08-558-735-9	Sequence 9, Appli
	30	30.4	3.8	3933	1	US-08-199-776-1	Sequence 1, Appli
c	31	30.4	3.8	901	2	US-08-508-448C-9	Sequence 9, Appli
	32	30.4	3.8	3933	5	PCT-US95-02044-1	Sequence 1, Appli
c	33	30.2	3.7	3143	2	US-08-485-621-1	Sequence 1, Appli
c	34	30.2	3.7	3143	4	US-08-973-831-1	Sequence 1, Appli
c	35	30.2	3.7	3143	5	PCT-US96-09530A-1	Sequence 1, Appli
	36	30	3.7	8438	1	US-07-945-283-1	Sequence 1, Appli
	37	29.8	3.7	996	4	US-08-924-759-21	Sequence 21, Appl
c	38	29.6	3.7	1186	4	US-08-731-722-5	Sequence 5, Appli
	39	29.4	3.6	1288	2	US-08-440-856A-9	Sequence 9, Appli
c	40	29.4	3.6	1517	2	US-08-508-448C-15	Sequence 15, Appl
	41	29.2	3.6	3480	1	US-07-657-769B-68	Sequence 68, Appl
c	42	29.2	3.6	17041	1	US-08-076-011-1	Sequence 1, Appli
c	43	29.2	3.6	30001	1	US-08-125-468-1	Sequence 1, Appli
	44	29.2	3.6	3480	1	US-07-789-184-219	Sequence 219, App
	45	29.2	3.6	3480	2	US-08-475-263-219	Sequence 219, App

ALIGNMENTS

```
RESULT 1
US-08-552-142A-16
; Sequence 16, Application US/08552142A
; Patent No. 5695995
;  GENERAL INFORMATION:
;    APPLICANT:  Weintraub, Harold M.
;    APPLICANT:  Lee, Jacqueline E.
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Hollenberg, Stanley M.
;    TITLE OF INVENTION:  Neurogenic Differentiation (NeuroD) Genes
;    TITLE OF INVENTION:  and Proteins
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/552,142A
;      FILING DATE:  02-NOV-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Broderick, Thomas F.
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  FHCR-1-8933
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-225-0709
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1462 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
;    IMMEDIATE SOURCE:
;      CLONE:  1.1.1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  231..1101
US-08-552-142A-16
```

```
  Query Match             4.5%;  Score 36.4;  DB 2;  Length 1462;
  Best Local Similarity   49.0%;  Pred. No. 0.042;
  Matches   97;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180
           | | ||||   |  || |  |  ||  || |  |||   | | |||  | | | ||
Db     261 CTCCTCTCGGACGTGCCCAAGTTCGCCAGCTGGGGCGACGGCGACGACGACGAGCCGAGG 320

Qy     181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
           | | | ||  | |||   ||||     | |  ||||||| | | | | | |  |  ||
Db     321 AGCGACAAGGGCGACGCGCCGCCGCAGCCTTCTCCTGCTCCCGGGTCGGGGGCTCCAGGA 380

Qy     241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300
           | ||   | | |     |   || ||  |   |  ||||     |    |||   ||||
Db     381 CCCGCCCGGGCCGCCAAGCCAGTGTCTCTTCGTGGAGGAGAAGAGATCCCTGAACCCACG 440

Qy     301 TCCGCCGGGGGCGAGGAG 318
           |  || | || | |||||
Db     441 TTGGCTGAGGTCAAGGAG 458


RESULT   2
US-08-910-973-16
; Sequence 16, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,973
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-10958
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1951 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
;    IMMEDIATE SOURCE:
;      CLONE:  1.1.1 (mouse neuroD2)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  230..1378
US-08-910-973-16


  Query Match             4.5%;  Score 36.4;  DB 2;  Length 1951;
  Best Local Similarity   49.0%;  Pred. No. 0.05;
  Matches   97;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180
           | | ||||   |  || |  |  ||  || |  |||   | | |||  | | | ||
Db     260 CTCCTCTCGGACGTGCCCAAGTTCGCCAGCTGGGGCGACGGCGACGACGACGAGCCGAGG 319

Qy     181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
           | | | ||  | |||   ||||     | |  ||||||| | | | | | |  |  ||
Db     320 AGCGACAAGGGCGACGCGCCGCCGCAGCCTTCTCCTGCTCCCGGGTCGGGGGCTCCAGGA 379

Qy     241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300
           | ||   | | |     |   || ||  |   |  ||||     |    |||   ||||
Db     380 CCCGCCCGGGCCGCCAAGCCAGTGTCTCTTCGTGGAGGAGAAGAGATCCCTGAACCCACG 439

Qy     301 TCCGCCGGGGGCGAGGAG 318
           |  || | || | |||||
Db     440 TTGGCTGAGGTCAAGGAG 457


RESULT   3
US-07-638-431-1/c
; Sequence 1, Application US/07638431
; Patent No. 5198535
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard
;    APPLICANT:  Khusmith, Srisin
;    APPLICANT:  Rogers IV, William O.
;    TITLE OF INVENTION:  Protective malaria sporozoite surface protein
;    TITLE OF INVENTION:  immunogen and gene
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  A. David Spevack
;      STREET:  NMRDC Building 1 T-12 National Naval
;      STREET:  Medical Center
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20814-5044
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/638,431
;      FILING DATE:  19910110
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, Avrom D.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-4033
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4673 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
;      STRAIN:  17X(NL)
;      DEVELOPMENTAL STAGE:  erythrocytic stage
;      TISSUE TYPE:  Blood
;      CELL TYPE:  erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
;      CLONE:  Py10.1111
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  718..3195
;      OTHER INFORMATION:
US-07-638-431-1


  Query Match             4.3%;  Score 34.6;  DB 1;  Length 4673;
  Best Local Similarity   55.4%;  Pred. No. 0.34;
  Matches   67;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     560 TCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTG 619
           | |  |||| |   |||  ||| | ||| ||    |     |||    |||| |    ||
Db    1993 TTGGCTTGTTTGGGTTTGATGGTTCGTTTGGGTTTGGCTTGTTTGGGTTTGGCTTGTTTG 1934

Qy     620 CTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATT 679
           ||| || |||||   |||| ||  | | |||||| |     |||||  |  |  |||  |
Db    1933 GTTTTGGCTTGTTTGGGTTTCGCCTTTTTGGGTTTCGCCTTTTTGGGTGGTTGTTTGGGT 1874

Qy     680 T 680
           |
Db    1873 T 1873


RESULT   4
PCT-US92-00018-1/c
; Sequence 1, Application PC/TUS9200018
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard
;    APPLICANT:  Khusmith, Srisin
;    APPLICANT:  Rogers IV, William O.
;    TITLE OF INVENTION:  Protective malaria sporozoite surface protein
;    TITLE OF INVENTION:  immunogen and gene encoding
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  A. David Spevack
;      STREET:  NMRDC Building 1 T-12 National Naval
;      STREET:  Medical Center
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20814-5044
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00018
;      FILING DATE:  19920103
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, Avram D.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-4033
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4673 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
;      STRAIN:  17X(NL)
;      DEVELOPMENTAL STAGE:  erythrocytic stage
;      TISSUE TYPE:  Blood
;      CELL TYPE:  erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
;      CLONE:  Py10.1111
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  718..3195
;      OTHER INFORMATION:
PCT-US92-00018-1


  Query Match             4.3%;  Score 34.6;  DB 5;  Length 4673;
  Best Local Similarity   55.4%;  Pred. No. 0.34;
  Matches   67;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     560 TCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTG 619
           | |  |||| |   |||  ||| | ||| ||    |     |||    |||| |    ||
Db    1993 TTGGCTTGTTTGGGTTTGATGGTTCGTTTGGGTTTGGCTTGTTTGGGTTTGGCTTGTTTG 1934

Qy     620 CTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATT 679
           ||| || |||||   |||| ||  | | |||||| |     |||||  |  |  |||  |
Db    1933 GTTTTGGCTTGTTTGGGTTTCGCCTTTTTGGGTTTCGCCTTTTTGGGTGGTTGTTTGGGT 1874

Qy     680 T 680
           |
Db    1873 T 1873


RESULT   5
5447867-2
;Patent No. 5447867
;    APPLICANT: BRIDGES, IAN;SCHUCH, WOLFGANG;GRIERSON, DONALD
;    TITLE OF INVENTION: RECOMBINANT DNA CONTAINING PECTIN
;ESTERASE GENE SEGEMENTS
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/24,866
;      FILING DATE: 26-FEB-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 720,629
;      FILING DATE: 25-JUN-1991
;      APPLICATION NUMBER: 419,779
;      FILING DATE: 29-SEP-1989
;      APPLICATION NUMBER: 119,614
;      FILING DATE: 12-NOV-1987
;SEQ ID NO:2:
;      LENGTH: 1636
5447867-2


  Query Match             4.1%;  Score 33.4;  DB 6;  Length 1636;
  Best Local Similarity   58.6%;  Pred. No. 0.43;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy     687 TATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGG 746
           |||||||  |  |||||  ||||| | | ||| | |  ||||  | |||  ||| ||| | |
Db     449 tatttcagtaaagatttttggatccttagaagcatctagtaaaatttcagactacaaaga 508

Qy     747 AACCAGACTTACAAGACTGAGTATTACTAATGAAACATT 785
           |  || |||   |       |  || |  | |||  ||
Db     509 tagaaggctttggattgcttttgatagtgttcaaaattt 547


RESULT   6
US-08-243-545-5/c
```

```
; Sequence 5, Application US/08243545
; Patent No. 5554512
; GENERAL INFORMATION:
;   APPLICANT:  Lyman, Stewart D.
;   APPLICANT:  Beckmann, M. Patricia
;   TITLE OF INVENTION:  Ligands for flt3/flk-2 Receptors
;   NUMBER OF SEQUENCES:  8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Stephen L. Malaska, Immunex Corporation
;     STREET:  51 University Street
;     CITY:  Seattle
;     STATE:  Washington
;     COUNTRY:  US
;     ZIP:  98101
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  Apple Macintosh
;     OPERATING SYSTEM:  Macintosh 7.0.1
;     SOFTWARE:  Microsoft Word, Version #5.1
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/243,545
;     FILING DATE:  11-MAY-1994
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/162,407
;     FILING DATE:  03-DEC-1993
;     APPLICATION NUMBER:  08/111,758
;     FILING DATE:  August 25, 1993
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/106,463
;     FILING DATE:  August 12, 1993
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/068,394
;     FILING DATE:  May 24, 1993
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Malaska, Stephen L.
;     REGISTRATION NUMBER:  32,655
;     REFERENCE/DOCKET NUMBER:  2813-C
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (206) 587-0430
;     TELEFAX:  (206) 233-0644
;     TELEX:  756822
; INFORMATION FOR SEQ ID NO:  5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  988 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA to mRNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  30..734
US-08-243-545-5


  Query Match             4.0%;  Score 32.8;  DB 1;  Length 988;
  Best Local Similarity   56.5%;  Pred. No. 0.49;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     127 TCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAG 186
           |||| |    ||| | ||  ||  | |   | | | || || |   ||| || ||||| |
Db     108 TCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGATAGGTTGTTGGGCTCCAGG 49

Qy     187 ATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACG 234
           ||||| | ||   | | | | | |||   || |||||   |  | ||
Db      48 CTGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAATTCCGGCCG 1
```

```
RESULT   7
US-08-539-304A-3
; Sequence 3, Application US/08539304A
; Patent No. 5792933
; GENERAL INFORMATION:
;   APPLICANT:  MA, DIN-POW
;   TITLE OF INVENTION:  FIBER-SPECIFIC PROTEIN EXPRESSION IN THE
;   TITLE OF INVENTION:  COTTON PLANT
;   NUMBER OF SEQUENCES:  7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;     STREET:  1755 JEFFERSON DAVIS HWY. SUITE 400
;     CITY:  ARLINGTON
;     STATE:  VA
;     COUNTRY:  USA
;     ZIP:  22202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/539,304A
;     FILING DATE:  04-OCT-1995
;     CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  NORMAN, OBLON F
;     REGISTRATION NUMBER:  24,618
;     REFERENCE/DOCKET NUMBER:  2343-037-27
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  703-413-3000
;     TELEFAX:  703-413-2220
; INFORMATION FOR SEQ ID NO:  3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1700 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   FEATURE:
;     NAME/KEY:  intron
;     LOCATION:  801..936
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  join(448..800, 937..943)
US-08-539-304A-3


  Query Match             4.0%;  Score 32.8;  DB 2;  Length 1700;
  Best Local Similarity   58.0%;  Pred. No. 0.69;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     631 TTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATT 690
           |||   |||   | | ||   | ||  | | ||   ||| ||||  |||| | | | |||
Db    1562 TTAGTATTTATGAGTATAATATATTTAGCATTTATTAGCATAATATATTTGAATTTGATT 1621

Qy     691 TCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACA 730
           |||   | || ||  ||||  | || || ||||| | |
Db    1622 AGAATTAGGTTTTTTCAACCTATAAATAGATGTAGTCAAA 1661


RESULT   8
US-08-993-962-5/c
; Sequence 5, Application US/08993962
; Patent No. 5843423
; GENERAL INFORMATION:
;   APPLICANT:  Lyman, Stewart D.
;   APPLICANT:  Beckmann, M. Patricia
;   TITLE OF INVENTION:  Ligands for flt3/flk-2 Receptors
;   NUMBER OF SEQUENCES:  8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Stephen L. Malaska, Immunex Corporation
;     STREET:  51 University Street
```

```
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0.1
;      SOFTWARE:  Microsoft Word, Version #5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,962
;      FILING DATE:  December 18, 1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,407
;      FILING DATE:  December 3, 1993
;      APPLICATION NUMBER:  08/111,758
;      FILING DATE:  August 25, 1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/106,463
;      FILING DATE:  August 12, 1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/068,394
;      FILING DATE:  May 24, 1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Malaska, Stephen L.
;      REGISTRATION NUMBER:  32,655
;      REFERENCE/DOCKET NUMBER:  2813-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;      TELEX:  756822
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  988 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  30..734
US-08-993-962-5

  Query Match             4.0%;  Score 32.8;  DB 3;  Length 988;
  Best Local Similarity   56.5%;  Pred. No. 0.49;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     127 TCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAG 186
           |||| |    ||| | ||  ||  | |   | | | || || |   ||| || ||||| |
Db     108 TCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGATAGGTTGTTGGGCTCCAGG 49

Qy     187 ATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACG 234
           ||||| | ||   | | | | | |||   || |||||   |  | ||
Db      48 CTGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAATTCCGGCCG 1


RESULT   9
PCT-US94-05365-5/c
; Sequence 5, Application PC/TUS9405365
;  GENERAL INFORMATION:
;    APPLICANT:  Lyman, Stewart D.
;    APPLICANT:  Beckmann, M. Patricia
;    TITLE OF INVENTION:  Ligands for flt3/flk-2 Receptors
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephen L. Malaska, Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05365
;      FILING DATE:  May 24, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  -to be assigned-
;      FILING DATE:  May 11, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/209,502
;      FILING DATE:  March 7, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/162,407
;      FILING DATE:  December 3, 1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/111,758
;      FILING DATE:  August 25, 1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/106,463
;      FILING DATE:  August 12, 1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/068,394
;      FILING DATE:  May 24, 1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Malaska, Stephen L.
;      REGISTRATION NUMBER:  32,655
;      REFERENCE/DOCKET NUMBER:  2813-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;      TELEX:  756822
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  988 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  30..734
PCT-US94-05365-5


  Query Match             4.0%;  Score 32.8;  DB 5;  Length 988;
  Best Local Similarity   56.5%;  Pred. No. 0.49;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     127 TCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAG 186
           |||| |    ||| | ||  ||  | |   | | | || || |   ||| || ||||| |
Db     108 TCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGATAGGTTGTTGGGCTCCAGG 49

Qy     187 ATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACG 234
           ||||| | ||   | | | | | |||   || |||||   |  | ||
Db      48 CTGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAATTCCGGCCG 1
```

```
RESULT  10
US-08-323-443B-1/c
; Sequence 1, Application US/08323443B
; Patent No. 5654170
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W.
;    APPLICANT:  LANDES, GREGORY M.
;    APPLICANT:  BURN, TIMOTHY C.
;    APPLICANT:  CONNORS, TIMOTHY D.
;    APPLICANT:  DACKOWSKI, WILLIAM R.
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,443B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  0372/0A462
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 527-7700
;      TELEFAX:  (212) 753-6237
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  PKD1 GENOMIC
US-08-323-443B-1


  Query Match             4.0%;  Score 32.4;  DB 1;  Length 31571;
  Best Local Similarity   50.5%;  Pred. No. 6.2;
  Matches  105;  Conservative    0;  Mismatches  101;  Indels    2;  Gaps    1;

Qy    108 CCTCGCCCTGTTGCCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGA 167
          ||| | ||||| || |    | ||  ||         |||  ||||  ||||     ||||
Db   3244 CCTGGGCCTGTGGCTCGGGGCGCTGGCGGGGGGCCCCGGGCGCGGCTGCGGGCCCTGCGA 3185

Qy    168 GAAAGTGTCGGTCTCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGC 227
          |      |   ||| |      ||| ||| |   ||| |||   ||||  || |   |||
Db   3184 GCCCCCCTGCCTCTGCGGCCCAGCGCCCGGCGCTCCGCCTGCCGCGTCAACTGCTCGGGC 3125

Qy    228 CGTGACGGCCAGA--CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGG 285
          || |     | ||  |||| |   |  ||| |  ||  | | | | | ||||  |  ||
Db   3124 CGCGGGCTGCGGACGCTCGGTCCCGCGCTGCGCATCCCCGCGGACGCCACAGCGCTGTGA 3065

Qy    286 GGGGCTGTTGCCACCTCCGCCGGGGGCG 313
          |  || |   |  |  |  ||||| | |
Db   3064 GTAGCGGGCCCAGCGGCACCCGGGAGAG 3037
```

```
RESULT  11
US-08-475-427-5/c
; Sequence 5, Application US/08475427
; Patent No. 5859340
;  GENERAL INFORMATION:
;    APPLICANT:  DUBOIS, Michel
;    APPLICANT:  GRISON, Rene
;    APPLICANT:  LEGUAY, Jean-Jacques
;    APPLICANT:  PIGNARD, Annie
;    APPLICANT:  TOPPAN, Alain
;    TITLE OF INVENTION:  Recombinant gene coding for a protein
;    TITLE OF INVENTION:  having endochitinase activity or for a precursor thereof.
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/475,427
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/842,165
;      FILING DATE:  01-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00607
;      FILING DATE:  21-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90 09460
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16781/564/BEDL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3012 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  2384..2462
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  2617..2697
;      OTHER INFORMATION:  /number= 2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1942..2383, 2463..2616, 2698..3007)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  1942..2166
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(2167..2383, 2463..2616, 2698..3007,
;      LOCATION:  2698)
US-08-475-427-5
```

```
  Query Match             4.0%;  Score 32.2;  DB 3;  Length 3012;
  Best Local Similarity   63.6%;  Pred. No. 1.6;
  Matches   49;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     717 TGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAA 776
           | ||| |   |  |||||  |||  |||| |  |||||||  |||| |   || |||  |
Db     920 TTGAAATTTATGAATTATGAATTTATATGTATTCAGACTTCAAAGATTTCTTAATACGGA 861

Qy     777 TGAAACATTTAGAAAAA 793
           | ||  |||| || |||
Db     860 TAAAGAATTTTGACAAA 844


RESULT  12
US-07-842-165-5/c
; Sequence 5, Application US/07842165
; Patent No. 5932698
;  GENERAL INFORMATION:
;    APPLICANT:  DUBOIS, Michel
;    APPLICANT:  GRISON, Rene
;    APPLICANT:  LEGUAY, Jean-Jacques
;    APPLICANT:  PIGNARD, Annie
;    APPLICANT:  TOPPAN, Alain
;    TITLE OF INVENTION:  Recombinant gene coding for a protein
;    TITLE OF INVENTION:  having endochitinase activity or for a precursor thereof.
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  King Street Station,Suite 500,1800 Diagonal
;      STREET:  Road,PO Box 299
;      CITY:  ALEXANDRIA
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/842,165
;      FILING DATE:  19920501
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAXE, Bernhard D
;      REGISTRATION NUMBER:  28,665
;      REFERENCE/DOCKET NUMBER:  16781/348
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-9300
;      TELEFAX:  (703) 683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3012 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  2384..2462
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  2617..2697
;      OTHER INFORMATION:  /number= 2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1942..2383, 2463..2616, 2698..3007)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  1942..2166
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(2167..2383, 2463..2616, 2698..3007, 2698)
US-07-842-165-5


  Query Match             4.0%;  Score 32.2;  DB 4;  Length 3012;
  Best Local Similarity   63.6%;  Pred. No. 1.6;
  Matches   49;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     717 TGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAA 776
           | ||| |   |  |||||  |||  |||| |  |||||||  |||| |   || |||  |
Db     920 TTGAAATTTATGAATTATGAATTTATATGTATTCAGACTTCAAAGATTTCTTAATACGGA 861

Qy     777 TGAAACATTTAGAAAAA 793
           | ||  |||| || |||
Db     860 TAAAGAATTTTGACAAA 844


RESULT  13
US-08-558-735-1/c
; Sequence 1, Application US/08558735
; Patent No. 5681724
;  GENERAL INFORMATION:
;    APPLICANT:  Tripp, Cynthia A.
;    APPLICANT:  Brandt, Kevin S.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  PARASITIC HELMINTH MACROPHAGE MIGRATION
;    TITLE OF INVENTION:  INHIBITORY FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/558,735
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-43
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  532 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..355
US-08-558-735-1


  Query Match             3.9%;  Score 31.6;  DB 1;  Length 532;
  Best Local Similarity   54.2%;  Pred. No. 0.8;
  Matches   64;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     390 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACAT 449
           ||||||  || | | || ||   ||||||| ||||  ||     ||||| |      |||
Db     208 AGAACCGATTGATTTTAAAACACACACAGCACATGGATCTGTACTTCCACCGAATGTCAT 149
```

```
Qy    450 AACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          |||||      | ||   | | ||| ||  |  || | ||      |  || ||||
Db    148 CGCTTGTCCACCATTCACATGGATTGATACGTAACTTTCCGGTTTTCCAAGTGCTTTT 91


RESULT 14
US-08-558-735-3/c
; Sequence 3, Application US/08558735
; Patent No. 5681724
;  GENERAL INFORMATION:
;    APPLICANT:  Tripp, Cynthia A.
;    APPLICANT:  Brandt, Kevin S.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  PARASITIC HELMINTH MACROPHAGE MIGRATION
;    TITLE OF INVENTION:  INHIBITORY FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/558,735
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-43
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-558-735-3


  Query Match             3.9%;  Score 31.6;  DB 1;  Length 345;
  Best Local Similarity   54.2%;  Pred. No. 0.61;
  Matches   64;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    390 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACAT 449
          ||||||  || | | || ||   ||||||| ||||  ||     ||||| |      |||
Db    201 AGAACCGATTGATTTTAAAACACACACAGCACATGGATCTGTACTTCCACCGAATGTCAT 142

Qy    450 AACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          |||||      | ||   | | ||| ||  |  || | ||      |  || ||||
Db    141 CGCTTGTCCACCATTCACATGGATTGATACGTAACTTTCCGGTTTTCCAAGTGCTTTT 84


RESULT 15
US-08-558-735-16/c
; Sequence 16, Application US/08558735
; Patent No. 5681724
;  GENERAL INFORMATION:
;    APPLICANT:  Tripp, Cynthia A.
;    APPLICANT:  Brandt, Kevin S.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  PARASITIC HELMINTH MACROPHAGE MIGRATION
;    TITLE OF INVENTION:  INHIBITORY FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/558,735
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-43
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  355 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-558-735-16


  Query Match             3.9%;  Score 31.6;  DB 1;  Length 355;
  Best Local Similarity   54.2%;  Pred. No. 0.62;
  Matches   64;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    390 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACAT 449
          ||||||  || | | || ||   ||||||| ||||  ||     ||||| |      |||
Db    208 AGAACCGATTGATTTTAAAACACACACAGCACATGGATCTGTACTTCCACCGAATGTCAT 149

Qy    450 AACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          |||||      | ||   | | ||| ||  |  || | ||      |  || ||||
Db    148 CGCTTGTCCACCATTCACATGGATTGATACGTAACTTTCCGGTTTTCCAAGTGCTTTT 91


Search completed: February 10, 2000, 13:25:30
Job time: 112 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 10, 2000, 17:55:13 ; Search time 23.51 Seconds
(without alignments)
147.013 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 269
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: OLIGO

Searched: 130275 seqs, 12848600 residues

Database : Issued_Patents_AA:*

Word size : 0

Number of hits that pass the threshold : 130275
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	2.6	452	1	US-08-290-978A-5	Sequence 5, Appli
2	7	2.6	302	1	US-08-121-054C-18	Sequence 18, Appl
3	7	2.6	302	1	US-08-121-054C-30	Sequence 30, Appl
4	7	2.6	211	1	US-08-276-852-34	Sequence 34, Appl
5	7	2.6	211	1	US-08-133-011-16	Sequence 16, Appl
6	7	2.6	482	1	US-08-358-160-5	Sequence 5, Appli
7	7	2.6	484	1	US-08-358-160-7	Sequence 7, Appli
8	7	2.6	399	1	US-08-530-950-10	Sequence 10, Appl
9	7	2.6	211	1	US-08-322-730A-16	Sequence 16, Appl
10	7	2.6	211	1	US-08-387-874-16	Sequence 16, Appl
11	7	2.6	211	1	US-08-899-575-34	Sequence 34, Appl
12	7	2.6	211	1	US-08-899-575-34	Sequence 34, Appl
13	7	2.6	1940	2	US-08-644-271-30	Sequence 30, Appl
14	7	2.6	461	2	US-08-463-587A-26	Sequence 26, Appl
15	7	2.6	452	2	US-08-780-869-5	Sequence 5, Appli
16	7	2.6	461	2	US-08-463-667A-4	Sequence 4, Appli
17	7	2.6	246	2	US-08-553-497A-24	Sequence 24, Appl
18	7	2.6	1186	2	US-08-861-464-8	Sequence 8, Appli
19	7	2.6	1186	2	US-08-396-001-8	Sequence 8, Appli
20	7	2.6	745	2	US-08-674-887A-8	Sequence 8, Appli
21	7	2.6	211	2	US-08-383-619-16	Sequence 16, Appl
22	7	2.6	233	2	US-08-829-876-101	Sequence 101, App
23	7	2.6	232	2	US-08-829-876-103	Sequence 103, App
24	7	2.6	233	2	US-08-829-876-105	Sequence 105, App
25	7	2.6	534	2	US-08-691-814B-8	Sequence 8, Appli
26	7	2.6	399	2	US-08-874-186-92	Sequence 92, Appl
27	7	2.6	461	3	PCT-US91-09133-27	Sequence 27, Appl
28	7	2.6	211	3	PCT-US93-08364-16	Sequence 16, Appl
29	7	2.6	293	3	PCT-US94-05669A-4	Sequence 4, Appli
30	7	2.6	293	3	PCT-US94-05669A-6	Sequence 6, Appli
31	7	2.6	349	3	PCT-US95-01185-139	Sequence 139, App
32	7	2.6	301	3	PCT-US95-01185-142	Sequence 142, App
33	7	2.6	335	3	PCT-US95-01185-143	Sequence 143, App
34	7	2.6	337	3	PCT-US95-01185-148	Sequence 148, App
35	7	2.6	349	3	PCT-US95-01185-151	Sequence 151, App
36	7	2.6	49	3	PCT-US95-01185-194	Sequence 194, App
37	7	2.6	60	3	PCT-US95-01185-195	Sequence 195, App
38	7	2.6	211	3	PCT-US95-08743-34	Sequence 34, Appl
39	7	2.6	2713	3	PCT-US96-01735-1	Sequence 1, Appli
40	6	2.2	907	1	US-07-718-575-2	Sequence 2, Appli
41	6	2.2	15	1	US-07-664-989B-11	Sequence 11, Appl
42	6	2.2	15	1	US-07-664-989B-18	Sequence 18, Appl
43	6	2.2	1385	1	US-07-876-280-2	Sequence 2, Appli
44	6	2.2	1385	1	US-07-675-772-2	Sequence 2, Appli
45	6	2.2	328	4	5212074-4	Patent No. 5212074

ALIGNMENTS

```
RESULT   1
US-08-290-978A-5
; Sequence 5, Application US/08290978A
; Patent No. 5624834
;  GENERAL INFORMATION:
;    APPLICANT:  KUSTERS-VAN SOMEREN, MARGO A.
;    APPLICANT:  MULLER, YVONNE
;    APPLICANT:  KESTER, HERMANUS C.M.
;    APPLICANT:  VISSER, JACOB
;    APPLICANT:  VAN OOYEN, ALBERT J.J.
;    APPLICANT:  ROLIN, CLAUS
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF THE
;    TITLE OF INVENTION:  EXO-POLYGALACTURONASE GENE FROM ASPERGILLUS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/290,978A
;      FILING DATE:  17-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  4615-0044.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-290-978A-5


  Query Match             2.6%;  Score 7;  DB 1;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGS 54
           |||||||
Db     336 GGGGSGS 342


RESULT   2
US-08-121-054C-18
```

```
; Sequence 18, Application US/08121054C
; Patent No. 5637481
;  GENERAL INFORMATION:
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Gilliland, Lisa K.
;    APPLICANT:  Hayden, Martha S.
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Bajorath, Jurgen
;    APPLICANT:  Fell, Perry
;    TITLE OF INVENTION:  Expression Vectors Encoding Bispecific
;    TITLE OF INVENTION:  Fusion Proteins and Methods of Producing Biologically
;    TITLE OF INVENTION:  Active Bispecific Fusion Proteins in a Mammalian Cell
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,054C
;      FILING DATE:  13-SEP-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/013,420
;      FILING DATE:  01-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436.18US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  302 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-121-054C-18

  Query Match             2.6%;  Score 7;  DB 1;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      47 LGGGGSG 53
           |||||||
Db     133 LGGGGSG 139


RESULT   3
US-08-121-054C-30
; Sequence 30, Application US/08121054C
; Patent No. 5637481
;  GENERAL INFORMATION:
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Gilliland, Lisa K.
;    APPLICANT:  Hayden, Martha S.
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Bajorath, Jurgen
;    APPLICANT:  Fell, Perry
;    TITLE OF INVENTION:  Expression Vectors Encoding Bispecific
;    TITLE OF INVENTION:  Fusion Proteins and Methods of Producing Biologically
;    TITLE OF INVENTION:  Active Bispecific Fusion Proteins in a Mammalian Cell
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,054C
;      FILING DATE:  13-SEP-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/013,420
;      FILING DATE:  01-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436.18US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  302 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-121-054C-30

  Query Match             2.6%;  Score 7;  DB 1;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      47 LGGGGSG 53
           |||||||
Db     133 LGGGGSG 139


RESULT   4
US-08-276-852-34
; Sequence 34, Application US/08276852
; Patent No. 5652138
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5652138th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,852
;      FILING DATE:  18-JUL-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
```

; APPLICATION NUMBER: US 08/178,302
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/954,148
; FILING DATE: 30-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCR1452P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 211 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; 8-276-852-34

Query Match 2.6%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GGGSGSG 55
|||||||
Db 55 GGGSGSG 61

RESULT 5
US-08-133-011-16
; Sequence 16, Application US/08133011
; Patent No. 5658727
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray
; APPLICANT: Barbas, Carlos
; APPLICANT: Lerner, Richard A.
; TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING
; TITLE OF INVENTION: PHAGEMIDS
; NUMBER OF SEQUENCES: 161
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5658727th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/133,011
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/683,602
; FILING DATE: 10-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/826,623
; FILING DATE: 27-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US 92/03091
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCRF 238.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 211 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
US-08-133-011-16

Query Match 2.6%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GGGSGSG 55
|||||||
Db 55 GGGSGSG 61

RESULT 6
US-08-358-160-5
; Sequence 5, Application US/08358160
; Patent No. 5663143
; GENERAL INFORMATION:
; APPLICANT: LEY, Arthur C.
; APPLICANT: LADNER, Robert C.
; APPLICANT: GUTERMAN, Sonia K.
; APPLICANT: ROBERTS, Bruce L.
; APPLICANT: MARKLAND, William
; APPLICANT: KENT, Rachel B.
; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
; NUMBER OF SEQUENCES: 234
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W. Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,160
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/133,031
; FILING DATE: 13-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/009,319
; FILING DATE: 26-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/664,989
; FILING DATE: 01-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/487,063
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/240,160
; FILING DATE: 02-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: LEY=1
; TELECOMMUNICATION INFORMATION:

```
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  482 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-358-160-5


  Query Match             2.6%;  Score 7;  DB 1;  Length 482;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db     326 GGGSGSG 332


RESULT   7
US-08-358-160-7
; Sequence 7, Application US/08358160
; Patent No. 5663143
;  GENERAL INFORMATION:
;    APPLICANT:  LEY, Arthur C.
;    APPLICANT:  LADNER, Robert C.
;    APPLICANT:  GUTERMAN, Sonia K.
;    APPLICANT:  ROBERTS, Bruce L.
;    APPLICANT:  MARKLAND, William
;    APPLICANT:  KENT, Rachel B.
;    TITLE OF INVENTION:  ENGINEERED HUMAN-DERIVED KUNITZ
;    TITLE OF INVENTION:  DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
;    NUMBER OF SEQUENCES:  234
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W. Suite 300
;      CITY:  Washington
;      STATE:  District of Columbia
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/358,160
;      FILING DATE:  16-DEC-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/133,031
;      FILING DATE:  13-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/009,319
;      FILING DATE:  26-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/664,989
;      FILING DATE:  01-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/487,063
;      FILING DATE:  02-MAR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/240,160
;      FILING DATE:  02-SEP-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cooper, Iver P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  LEY=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  484 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-358-160-7


  Query Match             2.6%;  Score 7;  DB 1;  Length 484;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db     328 GGGSGSG 334


RESULT   8
US-08-530-950-10
; Sequence 10, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  399 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  linear
US-08-530-950-10


  Query Match             2.6%;  Score 7;  DB 1;  Length 399;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
```

```
Db      13 GGGSGSG 19


RESULT   9
US-08-322-730A-16
; Sequence 16, Application US/08322730A
; Patent No. 5759817
;  GENERAL INFORMATION:
;    APPLICANT:  Barbas, Carlos
;    TITLE OF INVENTION:  HETERODIMERIC RECEPTOR LIBRARIES USING
;    TITLE OF INVENTION:  PHAGEMIDS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10550 No. 5759817th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/322,730A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/826,623
;      FILING DATE:  27-JAN-1992
;      APPLICATION NUMBER:  US 07/683,602
;      FILING DATE:  10-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bingham, Douglas A
;      REGISTRATION NUMBER:  32,457
;      REFERENCE/DOCKET NUMBER:  SCR0707P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-784-2937
;      TELEFAX:  619-784-9399
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  211 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
US-08-322-730A-16


  Query Match             2.6%;  Score 7;  DB 1;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db      55 GGGSGSG 61


RESULT  10
US-08-387-874-16
; Sequence 16, Application US/08387874
; Patent No. 5770356
;  GENERAL INFORMATION:
;    APPLICANT:  Light, Paul L., II
;    APPLICANT:  Lerner, Richard A.
;    TITLE OF INVENTION:  PHAGEMIDS COEXPRESSING A SURFACE
;    TITLE OF INVENTION:  RECEPTOR AND A SURFACE HETEROLOGOUS PROTEIN
;    NUMBER OF SEQUENCES:  97
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5770356th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/387,874
;      FILING DATE:  22-FEB-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08364
;      FILING DATE:  03-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/941,369
;      FILING DATE:  04-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  TSRI 303.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  211 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
US-08-387-874-16


  Query Match             2.6%;  Score 7;  DB 1;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db      55 GGGSGSG 61


RESULT  11
US-08-899-575-34
; Sequence 34, Application US/08899575
; Patent No. 5770440
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5770440th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,575
;      FILING DATE:  24-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/276,852
;      FILING DATE:  18-JUL-1994
;      APPLICATION NUMBER:  US 08/178,302
;      FILING DATE:  30-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  SCR1452P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  211 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
US-08-899-575-34


  Query Match             2.6%;  Score 7;  DB 1;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db      55 GGGSGSG 61


RESULT  12
US-08-899-575-34
; Sequence 34, Application US/08899575
; Patent No. 5804440
;  GENERAL INFORMATION:
;    APPLICANT:  Burton, Dennis R
;    APPLICANT:  Barbas, Carlos F
;    APPLICANT:  Lerner, Richard A
;    TITLE OF INVENTION:  HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;    TITLE OF INVENTION:  TO HUMAN IMMUNODEFICIENCY VIRUS
;    NUMBER OF SEQUENCES:  170
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 No. 5804440th Torrey Pines Road, Suite 220,
;      STREET:  Mail Drop TPC8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,575
;      FILING DATE:  24-JUL-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/276,852
;      FILING DATE:  18-JUL-1994
;      APPLICATION NUMBER:  US 08/178,302
;      FILING DATE:  30-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/954,148
;      FILING DATE:  30-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  SCR1452P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  211 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
US-08-899-575-34


  Query Match             2.6%;  Score 7;  DB 1;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db      55 GGGSGSG 61


RESULT  13
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1940 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Rat Agrin
;      LOCATION:  1...1940
;      OTHER INFORMATION:
US-08-644-271-30


  Query Match             2.6%;  Score 7;  DB 2;  Length 1940;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      50 GGSGSGE 56
           |||||||
Db     567 GGSGSGE 573


RESULT  14
US-08-463-587A-26
; Sequence 26, Application US/08463587A
; Patent No. 5821047
; GENERAL INFORMATION:
;    APPLICANT:  Garrard, Lisa J.
;    APPLICANT:  Henner, Dennis J.
;    APPLICANT:  Bass, Steven
;    APPLICANT:  Greene, Ronald
;    APPLICANT:  Lowman, Henry B.
;    APPLICANT:  Wells, James A.
;    APPLICANT:  Matthews, David J.
;    TITLE OF INVENTION:  ENRICHMENT METHOD FOR VARIANT PROTEINS WITH
;    TITLE OF INVENTION:  ALTERED BINDING PROPERTIES
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,587A
;      FILING DATE:  05-Jun-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/050058
;      FILING DATE:  30-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09133
;      FILING DATE:  03-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/743614
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/715300
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/683400
;      FILING DATE:  10-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/621667
;      FILING DATE:  03-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schwartz, Timothy R.
;      REGISTRATION NUMBER:  32171
;      REFERENCE/DOCKET NUMBER:  P0645P4D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-7467
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  461 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-463-587A-26


  Query Match             2.6%;  Score 7;  DB 2;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db     305 GGGSGSG 311


RESULT  15
US-08-780-869-5
; Sequence 5, Application US/08780869
; Patent No. 5830737
;  GENERAL INFORMATION:
;    APPLICANT:  KUSTERS-VAN SOMEREN, MARGO A.
;    APPLICANT:  MULLER, YVONNE
;    APPLICANT:  KESTER, HERMANUS C.M.
;    APPLICANT:  VISSER, JACOB
;    APPLICANT:  VAN OOYEN, ALBERT J.J.
;    APPLICANT:  ROLIN, CLAUS
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF THE
;    TITLE OF INVENTION:  EXO-POLYGALACTURONASE GENE FROM ASPERGILLUS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/780,869
;      FILING DATE:  24-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/290,978
;      FILING DATE:  17-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  4615-0044.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  452 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-780-869-5


  Query Match             2.6%;  Score 7;  DB 2;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGS 54
```

```
          !!!!!!!
Db     336 GGGGSGS 342

Search completed: February 10, 2000, 18:03:22
Job time: 489 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 22:00:30 ; Search time 42.58 Seconds
(without alignments)
2203.564 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: OLIGO_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17	2.1	3383	1	US-07-707-367-1	Sequence 1, Appli
	2	17	2.1	8438	1	US-07-945-283-1	Sequence 1, Appli
c	3	17	2.1	3306	1	US-08-261-206A-71	Sequence 71, Appl
c	4	17	2.1	1251	1	US-08-261-206A-75	Sequence 75, Appl
	5	17	2.1	2625	3	US-08-759-945-1	Sequence 1, Appli
	6	17	2.1	1299	3	US-08-994-719C-3	Sequence 3, Appli
c	7	16	2.0	1393	1	US-08-052-205-2	Sequence 2, Appli
c	8	16	2.0	1470	1	US-08-052-205-3	Sequence 3, Appli
	9	16	2.0	1451	1	US-08-031-143B-68	Sequence 68, Appl
	10	16	2.0	301	1	US-08-031-143B-76	Sequence 76, Appl
c	11	16	2.0	2542	1	US-08-120-960-1	Sequence 1, Appli
	12	16	2.0	966	1	US-08-162-475A-1	Sequence 1, Appli
	13	16	2.0	935	1	US-08-162-475A-3	Sequence 3, Appli
c	14	16	2.0	1393	2	US-08-595-974-2	Sequence 2, Appli
c	15	16	2.0	1470	2	US-08-595-974-3	Sequence 3, Appli
	16	16	2.0	6453	2	US-08-306-691B-14	Sequence 14, Appl
	17	16	2.0	26	2	US-08-244-468-3	Sequence 3, Appli
	18	16	2.0	1473	2	US-08-672-571A-2	Sequence 2, Appli
	19	16	2.0	1386	2	US-08-672-571A-4	Sequence 4, Appli
c	20	16	2.0	10898	4	US-08-481-658B-5	Sequence 5, Appli
	21	16	2.0	3680	4	US-08-494-907-1	Sequence 1, Appli
	22	16	2.0	5076	4	US-08-494-907-2	Sequence 2, Appli
	23	16	2.0	6387	4	US-08-494-907-3	Sequence 3, Appli
	24	16	2.0	6170	4	US-08-494-907-4	Sequence 4, Appli
	25	16	2.0	1047	4	US-08-494-907-11	Sequence 11, Appl
c	26	16	2.0	10898	4	US-08-477-504A-5	Sequence 5, Appli
c	27	16	2.0	10898	4	US-08-486-756A-5	Sequence 5, Appli
c	28	16	2.0	10898	4	US-08-485-862B-5	Sequence 5, Appli
	29	16	2.0	3680	5	PCT-US96-10986-1	Sequence 1, Appli
	30	16	2.0	5076	5	PCT-US96-10986-2	Sequence 2, Appli
	31	16	2.0	6387	5	PCT-US96-10986-3	Sequence 3, Appli
	32	16	2.0	6170	5	PCT-US96-10986-4	Sequence 4, Appli
	33	16	2.0	1047	5	PCT-US96-10986-11	Sequence 11, Appl
c	34	16	2.0	1730	6	5223391-8	Patent No. 5223391
	35	15	1.9	2516	1	US-07-914-282D-2	Sequence 2, Appli
c	36	15	1.9	4380	1	US-07-582-945-1	Sequence 1, Appli
c	37	15	1.9	2608	1	US-07-725-083-1	Sequence 1, Appli
	38	15	1.9	2051	1	US-07-864-475A-3	Sequence 3, Appli
	39	15	1.9	2516	1	US-08-276-887A-2	Sequence 2, Appli
c	40	15	1.9	894	1	US-08-178-708-7	Sequence 7, Appli
	41	15	1.9	3987	1	US-07-688-352C-19	Sequence 19, Appl
	42	15	1.9	78	1	US-08-281-702A-5	Sequence 5, Appli
c	43	15	1.9	9171	1	US-08-038-682-5	Sequence 5, Appli
c	44	15	1.9	9323	1	US-08-038-682-6	Sequence 6, Appli
	45	15	1.9	2820	5	PCT-US94-05905-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-07-707-367-1
; Sequence 1, Application US/07707367
; Patent No. 5196316
; GENERAL INFORMATION:
; APPLICANT: Iwasaki, Yasuno
; APPLICANT: Shimoi, Hiroko
; APPLICANT: Suzuki, Kenji
; APPLICANT: Ghisalba, Oreste
; APPLICANT: Nishikawa, Yoshiki
; APPLICANT: Kawahara, Takashi
; APPLICANT: Kangawa, Kenji
; TITLE OF INVENTION: No. 5196316el Enzyme and DNA Coding Therefor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/707,367
; FILING DATE: 19910530
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 141678/90
; FILING DATE: 01-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 210535/90
; FILING DATE: 10-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 329911/90
; FILING DATE: 30-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-18110/A/CGJ 44
; TELECOMMUNICATION INFORMATION: /
; TELEPHONE: (914)785-7120
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3383 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (recombinant)
; ORIGINAL SOURCE:
; ORGANISM: Xenopus laevis
; INDIVIDUAL ISOLATE: DNA encoding protein AE-III,

; INDIVIDUAL ISOLATE: precursor to PHL enzyme
; IMMEDIATE SOURCE:
; CLONE: pAE-III-202-4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..2835
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1177..2145
US-07-707-367-1

Query Match 2.1%; Score 17; DB 1; Length 3383;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 335 ACCTCAAAGTGGGACAA 351
|||||||||||||||||
1241 ACCTCAAAGTGGGACAA 1257

RESULT 2
US-07-945-283-1
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1267, "t")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
US-07-945-283-1

Query Match 2.1%; Score 17; DB 1; Length 8438;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 CCAAGATGGCGGCCGCC 198
|||||||||||||||||
Db 7222 CCAAGATGGCGGCCGCC 7238

RESULT 3
US-08-261-206A-71/c
; Sequence 71, Application US/08261206A
; Patent No. 5574007
; GENERAL INFORMATION:
; APPLICANT: Zushi, Mitichitaka
; APPLICANT: Gomi, Komakazu
; APPLICANT: Yamamoto, Shuji
; APPLICANT: Suzuki, Koji
; APPLICANT: Matsuda, Akio
; TITLE OF INVENTION: A Polypeptide Capable of Interacting
; TITLE OF INVENTION: with Thrombin
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 N. Washington St.
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22046-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,206A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/740,492
; FILING DATE: 03-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30330
; REFERENCE/DOCKET NUMBER: 216-275P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Acremonium chrysogenum
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3306

```
;      OTHER INFORMATION:  /label= PGK_gene
;      OTHER INFORMATION:  /note= "Nucleotide sequence of region A in Figure
;      OTHER INFORMATION:  59.  The sequence is presented as Figure 61."
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1252..1317
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1463..1883
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1948..2715
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1252..1317, 1463..1883, 1948..2714)
US-08-261-206A-71


  Query Match             2.1%;  Score 17;  DB 1;  Length 3306;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    126 CTCCCTCCCGCTCCTGG 142
          |||||||||||||||||
Db    209 CTCCCTCCCGCTCCTGG 193


RESULT   4
US-08-261-206A-75/c
; Sequence 75, Application US/08261206A
; Patent No. 5574007
;  GENERAL INFORMATION:
;    APPLICANT:  Zushi, Mitichitaka
;    APPLICANT:  Gomi, Komakazu
;    APPLICANT:  Yamamoto, Shuji
;    APPLICANT:  Suzuki, Koji
;    APPLICANT:  Matsuda, Akio
;    TITLE OF INVENTION:  A Polypeptide Capable of Interacting
;    TITLE OF INVENTION:  with Thrombin
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 N. Washington St.
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22046-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/261,206A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/740,492
;      FILING DATE:  03-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svensson, Leonard R.
;      REGISTRATION NUMBER:  30330
;      REFERENCE/DOCKET NUMBER:  216-275P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  75:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1251 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Acremonium chrysogenum
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..1251
;      OTHER INFORMATION:  /label= gene_sequence
;      OTHER INFORMATION:  /note= "5' untranslated region of PGK gene from A.
;      OTHER INFORMATION:  chrysogenum."
US-08-261-206A-75


  Query Match             2.1%;  Score 17;  DB 1;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    126 CTCCCTCCCGCTCCTGG 142
          |||||||||||||||||
Db    209 CTCCCTCCCGCTCCTGG 193


RESULT   5
US-08-759-945-1
; Sequence 1, Application US/08759945
; Patent No. 5834249
;  GENERAL INFORMATION:
;    APPLICANT:  KAZUAKI, Furukawa
;    APPLICANT:  SUGIMURA, Keijiro
;    APPLICANT:  OHSUYE, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/759,945
;      FILING DATE:  03-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/303,191
;      FILING DATE:  08-SEP-1994
;      APPLICATION NUMBER:  JP 5-257881
;      FILING DATE:  08-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa Stanek
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  001560-227
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2625 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2625
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  118..2625
US-08-759-945-1
```

```
  Query Match             2.1%;  Score 17;  DB 3;  Length 2625;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    335 ACCTCAAAGTGGGACAA 351
          |||||||||||||||||
Db   1223 ACCTCAAAGTGGGACAA 1239


RESULT   6
US-08-994-719C-3
; Sequence 3, Application US/08994719C
; Patent No. 5908626
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Liming; Chang, Tse Wen
;    TITLE OF INVENTION:  Hybrid with Interferon-  and an
;    TITLE OF INVENTION:  Immunoglobulin Fc Joined by a Peptide Liner
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Tanox Biosystems, Inc.
;      STREET:  10301 Stella Link Rd.
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  Addonics C142 SVGA
;      OPERATING SYSTEM:  DOS 3.30
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,719C
;      FILING DATE:  12/19/1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/719,331
;      FILING DATE:  09/25/1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mirabel, Eric P.
;      REGISTRATION NUMBER:  31,211
;      REFERENCE/DOCKET NUMBER:  TNX95-2AA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 664-2288
;      TELEFAX:  (713) 664-8914
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1299 nucleic acids
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
US-08-994-719C-3


  Query Match             2.1%;  Score 17;  DB 3;  Length 1299;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    142 GGCGGAGGCGGAAGCGG 158
          |||||||||||||||||
Db    580 GGCGGAGGCGGAAGCGG 596


RESULT   7
US-08-052-205-2/c
; Sequence 2, Application US/08052205
; Patent No. 5510259
;  GENERAL INFORMATION:
;    APPLICANT:  SUGAMURA, KAZUO
;    APPLICANT:  TAKESHITA, TOSHIKAZU
;    APPLICANT:  ASAO, HIRONOBU
;    APPLICANT:  NAKAMURA, MASATAKA
;    APPLICANT:  SHIMAMURA, TOSHIRO
;    APPLICANT:  SUZUKI, MANABU
;    APPLICANT:  HAMURO, JUNJI
;    TITLE OF INVENTION:  HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/052,205
;      FILING DATE:  19930422
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 104947/1992
;      FILING DATE:  23-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5510259man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-615-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-052-205-2


  Query Match             2.0%;  Score 16;  DB 1;  Length 1393;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    272 CTACAGGACCCTGGGG 287
          ||||||||||||||||
Db   1086 CTACAGGACCCTGGGG 1071


RESULT   8
US-08-052-205-3/c
; Sequence 3, Application US/08052205
; Patent No. 5510259
;  GENERAL INFORMATION:
;    APPLICANT:  SUGAMURA, KAZUO
;    APPLICANT:  TAKESHITA, TOSHIKAZU
;    APPLICANT:  ASAO, HIRONOBU
;    APPLICANT:  NAKAMURA, MASATAKA
;    APPLICANT:  SHIMAMURA, TOSHIRO
;    APPLICANT:  SUZUKI, MANABU
;    APPLICANT:  HAMURO, JUNJI
;    TITLE OF INVENTION:  HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
```

```
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/052,205
;      FILING DATE:  19930422
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 104947/1992
;      FILING DATE:  23-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5510259man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-615-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1470 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  15..1121
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  81..1121
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  15..80
US-08-052-205-3


  Query Match             2.0%;  Score 16;  DB 1;  Length 1470;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 CTACAGGACCCTGGGG 287
           ||||||||||||||||
Db    1163 CTACAGGACCCTGGGG 1148


RESULT   9
US-08-031-143B-68/c
; Sequence 68, Application US/08031143B
; Patent No. 5518880
;  GENERAL INFORMATION:
;    APPLICANT:  LEONARD, WARREN J.; NOGUCHI, MASAYUKI;
;    APPLICANT:  MCBRIDE, O. WESLEY
;    TITLE OF INVENTION:  METHODS FOR DIAGNOSIS AND
;    TITLE OF INVENTION:  TREATMENT OF XSCID
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVE.
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORD PERFECT # 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/031,143B
;      FILING DATE:  12-MAR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAM S. FEILER
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-4800
;      TELEFAX:  212-751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  68:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1451
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  UNKNOWN
;    MOLECULE TYPE:
;      DESCRIPTION:  OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
;      INDIVIDUAL ISOLATE:  IL-2R
US-08-031-143B-68


  Query Match             2.0%;  Score 16;  DB 1;  Length 1451;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 CTACAGGACCCTGGGG 287
           ||||||||||||||||
Db    1163 CTACAGGACCCTGGGG 1148


RESULT  10
US-08-031-143B-76/c
; Sequence 76, Application US/08031143B
; Patent No. 5518880
;  GENERAL INFORMATION:
;    APPLICANT:  LEONARD, WARREN J.; NOGUCHI, MASAYUKI;
;    APPLICANT:  MCBRIDE, O. WESLEY
;    TITLE OF INVENTION:  METHODS FOR DIAGNOSIS AND
;    TITLE OF INVENTION:  TREATMENT OF XSCID
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVE.
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORD PERFECT # 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/031,143B
;      FILING DATE:  12-MAR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAM S. FEILER
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-4800
;      TELEFAX:  212-751-6849
;      TELEX:  421792
```

```
;  INFORMATION FOR SEQ ID NO:  76:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  301
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  UNKNOWN
;    MOLECULE TYPE:
;      DESCRIPTION:  OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
;      INDIVIDUAL ISOLATE:  IL-2R
US-08-031-143B-76


  Query Match             2.0%;  Score 16;  DB 1;  Length 301;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    272 CTACAGGACCCTGGGG 287
          ||||||||||||||||
Db    301 CTACAGGACCCTGGGG 286


RESULT  11
US-08-120-960-1/c
; Sequence 1, Application US/08120960
; Patent No. 5523225
;  GENERAL INFORMATION:
;    APPLICANT:  KRAUS, JAN P
;    TITLE OF INVENTION:  DNA SEQUENCE ENCODING HUMAN
;    TITLE OF INVENTION:  CYSTATHIONINE B-SYNTHASE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DILWORTH & BARRESE
;      STREET:  4350 LA JOLLA VILLAGE DRIVE, SUITE 300
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  U.S.A.
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/120,960
;      FILING DATE:  12-SEP-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PEPPER PH.D., FREDERICK W.
;      REGISTRATION NUMBER:  31,286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4410
;      TELEFAX:  619-453-2839
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2542 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  181..1834
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  181..1834
US-08-120-960-1


  Query Match             2.0%;  Score 16;  DB 1;  Length 2542;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    131 TCCCGCTCCTGGGCGG 146
          ||||||||||||||||
Db   1826 TCCCGCTCCTGGGCGG 1811


RESULT  12
US-08-162-475A-1
; Sequence 1, Application US/08162475A
; Patent No. 5656474
;  GENERAL INFORMATION:
;    APPLICANT:  Zohreh Tabaeizadeh
;    TITLE OF INVENTION:  A novel endochitinase gene
;    TITLE OF INVENTION:  induced by osmotic stress and abscisic acid isolated
;    TITLE OF INVENTION:  from the wild tomato Lycopersicon chilense Dun.
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 New York Avenue, N.W., 9th Floor
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 720 kb diskette
;      COMPUTER:  IBM PS/2, Model 30
;      OPERATING SYSTEM:  PC-DOS 3.30
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,475A
;      FILING DATE:  December 7, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, Paul E. Jr.
;      REGISTRATION NUMBER:  32011
;      REFERENCE/DOCKET NUMBER:  PEW/3122/204351
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  966 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:  Lycopersicon chilense
US-08-162-475A-1


  Query Match             2.0%;  Score 16;  DB 1;  Length 966;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    565 TTGTCTCTTTTTCTTG 580
          ||||||||||||||||
Db     56 TTGTCTCTTTTTCTTG 71


RESULT  13
US-08-162-475A-3
; Sequence 3, Application US/08162475A
; Patent No. 5656474
;  GENERAL INFORMATION:
;    APPLICANT:  Zohreh Tabaeizadeh
;    TITLE OF INVENTION:  A novel endochitinase gene
;    TITLE OF INVENTION:  induced by osmotic stress and abscisic acid isolated
;    TITLE OF INVENTION:  from the wild tomato Lycopersicon chilense Dun.
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
```

```
;       STREET:  1100 New York Avenue, N.W., 9th Floor
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005-3918
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 720 kb diskette
;       COMPUTER:  IBM PS/2, Model 30
;       OPERATING SYSTEM:  PC-DOS 3.30
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,475A
;       FILING DATE:  December 7, 1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, Paul E. Jr.
;       REGISTRATION NUMBER:  32011
;       REFERENCE/DOCKET NUMBER:  PEW/3122/204351
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  935 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
;     MOLECULE TYPE:  cDNA to mRNA
;     ORIGINAL SOURCE:  Lycopersicon chilense
US-08-162-475A-3


  Query Match             2.0%;  Score 16;  DB 1;  Length 935;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     565 TTGTCTCTTTTTCTTG 580
           ||||||||||||||||
Db      27 TTGTCTCTTTTTCTTG 42


RESULT  14
US-08-595-974-2/c
; Sequence 2, Application US/08595974
; Patent No. 5705608
;  GENERAL INFORMATION:
;    APPLICANT:  SUGAMURA, KAZUO
;    APPLICANT:  TAKESHITA, TOSHIKAZU
;    APPLICANT:  ASAO, HIRONOBU
;    APPLICANT:  NAKAMURA, MASATAKA
;    APPLICANT:  SHIMAMURA, TOSHIRO
;    APPLICANT:  SUZUKI, MANABU
;    APPLICANT:  HAMURO, JUNJI
;    TITLE OF INVENTION:  HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/595,974
;      FILING DATE:  06-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/052,205
;      FILING DATE:  22-APR-1993
;      APPLICATION NUMBER:  JP 104947/1992
;      FILING DATE:  23-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5705608man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-615-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-595-974-2


  Query Match             2.0%;  Score 16;  DB 2;  Length 1393;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     272 CTACAGGACCCTGGGG 287
           ||||||||||||||||
Db    1086 CTACAGGACCCTGGGG 1071


RESULT  15
US-08-595-974-3/c
; Sequence 3, Application US/08595974
; Patent No. 5705608
;  GENERAL INFORMATION:
;    APPLICANT:  SUGAMURA, KAZUO
;    APPLICANT:  TAKESHITA, TOSHIKAZU
;    APPLICANT:  ASAO, HIRONOBU
;    APPLICANT:  NAKAMURA, MASATAKA
;    APPLICANT:  SHIMAMURA, TOSHIRO
;    APPLICANT:  SUZUKI, MANABU
;    APPLICANT:  HAMURO, JUNJI
;    TITLE OF INVENTION:  HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/595,974
;      FILING DATE:  06-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/052,205
;      FILING DATE:  22-APR-1993
;      APPLICATION NUMBER:  JP 104947/1992
;      FILING DATE:  23-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5705608man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-615-0X
```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 15..1121
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 81..1121
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 15..80
US-08-595-974-3

Query Match 2.0%; Score 16; DB 2; Length 1470;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    272 CTACAGGACCCTGGGG 287
          ||||||||||||||||
Db   1163 CTACAGGACCCTGGGG 1148
```

Search completed: February 10, 2000, 22:02:56
Job time: 146 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 00:39:47 ; Search time 530.22 Seconds
(without alignments)
-4638.586 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold : 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_prl:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	64.6	8.0	149592	33	AC005718	AC005718 Drosophil
c	2	64.6	8.0	188633	35	AC007175	AC007175 Drosophil
	3	64.6	8.0	83204	46	AC004434	Ac004434 Drosophil
	4	64.6	8.0	132637	46	AC006092	Ac006092 Drosophil
c	5	49.6	6.1	40912	35	CELC41D11	AF003740 Caenorhab
	6	47.2	5.8	205157	44	AC016672	AC016672 Homo sapi
	7	46.2	5.7	117674	32	DMBR30C13	AL122025 Drosophil
	8	46.2	5.7	1440	35	AF181623	AF181623 Drosophil
	9	43.2	5.3	22333	34	CELC02F5	L14745 C. elegans
c	10	42.8	5.3	49649	16	CVY15035	Y15035 Cowpox viru
c	11	42.6	5.3	18753	2	AE001582	AE001582 Borrelia
c	12	41.6	5.1	88128	7	AB017069	AB017069 Arabidops
c	13	41.6	5.1	73384	40	AF178650	AF178650 Homo sapi
	14	39.8	4.9	135301	16	BHV1CGEN	AJ004801 Bovine he
c	15	39.8	4.9	135301	16	BHV1CGEN	AJ004801 Bovine he
c	16	39.8	4.9	8113	16	HSBBICP4A	L14320 Bovine herp
c	17	39.8	4.9	95456	33	AC002092	AC002092 Homo sapi
	18	39.8	4.9	93713	40	AC006071	AC006071 Homo sapi
	19	39.6	4.9	148051	32	AP000726	AP000726 Homo sapi
c	20	39.2	4.8	134977	11	AC002347	AC002347 Homo sapi
c	21	39	4.8	204929	45	AC009362	AC009362 Homo sapi
	22	38.8	4.8	49173	42	AC011322	AC011322 Homo sapi
c	23	38.6	4.8	85835	8	ATAC005956	AC005956 Arabidops
	24	38.6	4.8	158147	44	AC007926	AC007926 Trypanoso
	25	38.4	4.7	123737	41	AF189001	AF189001 Homo sapi
c	26	38.2	4.7	219917	33	AC007612	AC007612 Homo sapi
	27	38	4.7	1831	12	MUSNDRF	D83507 Mouse mRNA
	28	38	4.7	12971	35	AF088979	AF088979 Dictyoste
	29	38	4.7	179169	44	AC011946	AC011946 Homo sapi
	30	37.8	4.7	14548	2	AF108766	AF108766 Rhodobact
	31	37.8	4.7	126581	11	AC003957	AC003957 Homo sapi
c	32	37.8	4.7	35414	11	AC005328	AC005328 Homo sapi
c	33	37.8	4.7	43514	11	AC005545	AC005545 Homo sapi
c	34	37.8	4.7	170820	41	AC006286	AC006286 Homo sapi
	35	37.8	4.7	194841	42	AC008155	AC008155 Homo sapi
	36	37.8	4.7	165565	42	AC010775	AC010775 Homo sapi
c	37	37.6	4.6	100000	10	AP000502	AP000502 Homo sapi
	38	37.6	4.6	210134	10	CNS018OV	AL109758 Human chr
c	39	37.4	4.6	126198	11	AC003089	AC003089 Human BAC
c	40	37.4	4.6	86894	11	AC003676	AC003676 Homo sapi
	41	37.4	4.6	1582	40	AF151867	AF151867 Homo sapi
	42	37.4	4.6	171681	40	HUAC002299	AC002299 Homo sapi
	43	37.4	4.6	176384	42	AC009608	AC009608 Homo sapi
c	44	37.4	4.6	141193	44	AC015721	AC015721 Homo sapi
	45	37.2	4.6	108803	11	HS550H1	AL035420 Human DNA

ALIGNMENTS

RESULT 1
AC005718/c
LOCUS AC005718 149592 bp DNA HTG 30-JUL-1999
DEFINITION Drosophila melanogaster chromosome 2 clone DS02336 (D440) map 60C8-60D2 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 68 unordered pieces.
ACCESSION AC005718
VERSION AC005718.10 GI:5656710
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 149592)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Sequencing of Drosophila melanogaster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149592)
AUTHORS Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Jul 30, 1999 this sequence version replaced gi:5630036.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases. P1 library location: 25-32.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	343: contig of 343 bp in length
*	344	423: gap of unknown length
*	424	1169: contig of 746 bp in length
*	1170	1249: gap of unknown length
*	1250	2094: contig of 845 bp in length
*	2095	2174: gap of unknown length
*	2175	3405: contig of 1231 bp in length
*	3406	3485: gap of unknown length
*	3486	4245: contig of 760 bp in length
*	4246	4325: gap of unknown length
*	4326	5054: contig of 729 bp in length
*	5055	5134: gap of unknown length
*	5135	6031: contig of 897 bp in length
*	6032	6111: gap of unknown length
*	6112	7163: contig of 1052 bp in length
*	7164	7243: gap of unknown length
*	7244	8013: contig of 770 bp in length
*	8014	8093: gap of unknown length
*	8094	8951: contig of 858 bp in length
*	8952	9031: gap of unknown length
*	9032	9815: contig of 784 bp in length
*	9816	9895: gap of unknown length
*	9896	10572: contig of 677 bp in length
*	10573	10652: gap of unknown length
*	10653	11442: contig of 790 bp in length
*	11443	11522: gap of unknown length
*	11523	12232: contig of 710 bp in length
*	12233	12312: gap of unknown length
*	12313	13278: contig of 966 bp in length
*	13279	13358: gap of unknown length
*	13359	14011: contig of 653 bp in length
*	14012	14091: gap of unknown length
*	14092	14897: contig of 806 bp in length
*	14898	14977: gap of unknown length
*	14978	15740: contig of 763 bp in length
*	15741	15820: gap of unknown length
*	15821	16793: contig of 973 bp in length
*	16794	16873: gap of unknown length
*	16874	17825: contig of 952 bp in length
*	17826	17905: gap of unknown length
*	17906	18911: contig of 1006 bp in length
*	18912	18991: gap of unknown length
*	18992	19720: contig of 729 bp in length
*	19721	19800: gap of unknown length
*	19801	20785: contig of 985 bp in length
*	20786	20865: gap of unknown length
*	20866	22392: contig of 1527 bp in length
*	22393	22472: gap of unknown length
*	22473	24290: contig of 1818 bp in length
*	24291	24370: gap of unknown length
*	24371	25694: contig of 1324 bp in length
*	25695	25774: gap of unknown length
*	25775	26731: contig of 957 bp in length
*	26732	26811: gap of unknown length
*	26812	27793: contig of 982 bp in length
*	27794	27873: gap of unknown length
*	27874	29161: contig of 1288 bp in length
*	29162	29241: gap of unknown length
*	29242	30491: contig of 1250 bp in length
*	30492	30571: gap of unknown length
*	30572	32263: contig of 1692 bp in length
*	32264	32343: gap of unknown length
*	32344	34053: contig of 1710 bp in length
*	34054	34133: gap of unknown length
*	34134	35477: contig of 1344 bp in length
*	35478	35557: gap of unknown length
*	35558	36683: contig of 1126 bp in length
*	36684	36763: gap of unknown length
*	36764	39017: contig of 2254 bp in length
*	39018	39097: gap of unknown length
*	39098	41095: contig of 1998 bp in length
*	41096	41175: gap of unknown length
*	41176	42262: contig of 1087 bp in length
*	42263	42342: gap of unknown length
*	42343	43644: contig of 1302 bp in length
*	43645	43724: gap of unknown length
*	43725	45000: contig of 1276 bp in length
*	45001	45080: gap of unknown length
*	45081	46626: contig of 1546 bp in length
*	46627	46706: gap of unknown length
*	46707	50898: contig of 4192 bp in length
*	50899	50978: gap of unknown length
*	50979	54635: contig of 3657 bp in length
*	54636	54715: gap of unknown length
*	54716	59775: contig of 5060 bp in length
*	59776	59855: gap of unknown length
*	59856	70727: contig of 10872 bp in length
*	70728	70807: gap of unknown length
*	70808	132766: contig of 61959 bp in length
*	132767	132846: gap of unknown length
*	132847	133588: contig of 742 bp in length
*	133589	133668: gap of unknown length
*	133669	134331: contig of 663 bp in length
*	134332	134411: gap of unknown length
*	134412	135076: contig of 665 bp in length
*	135077	135156: gap of unknown length
*	135157	135778: contig of 622 bp in length
*	135779	135858: gap of unknown length
*	135859	136505: contig of 647 bp in length
*	136506	136585: gap of unknown length
*	136586	137259: contig of 674 bp in length
*	137260	137339: gap of unknown length
*	137340	137991: contig of 652 bp in length
*	137992	138071: gap of unknown length
*	138072	138762: contig of 691 bp in length

```
            *   138763   138842: gap of unknown length
            *   138843   139393: contig of 551 bp in length
            *   139394   139473: gap of unknown length
            *   139474   140102: contig of 629 bp in length
            *   140103   140182: gap of unknown length
            *   140183   140811: contig of 629 bp in length
            *   140812   140891: gap of unknown length
            *   140892   141500: contig of 609 bp in length
            *   141501   141580: gap of unknown length
            *   141581   142262: contig of 682 bp in length
            *   142263   142342: gap of unknown length
            *   142343   143046: contig of 704 bp in length
            *   143047   143126: gap of unknown length
            *   143127   143812: contig of 686 bp in length
            *   143813   143892: gap of unknown length
            *   143893   144614: contig of 722 bp in length
            *   144615   144694: gap of unknown length
            *   144695   145373: contig of 679 bp in length
            *   145374   145453: gap of unknown length
            *   145454   146116: contig of 663 bp in length
            *   146117   146196: gap of unknown length
            *   146197   146808: contig of 612 bp in length
            *   146809   146888: gap of unknown length
            *   146889   147514: contig of 626 bp in length
            *   147515   147594: gap of unknown length
            *   147595   148161: contig of 567 bp in length
            *   148162   148241: gap of unknown length
            *   148242   148927: contig of 686 bp in length
            *   148928   149007: gap of unknown length
            *   149008   149592: contig of 585 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149592
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /clone="DS02336 (D440)"
                     /chromosome="2"
                     /clone_lib="Pl library, partial Sau3A in pNS582tet14Ad10"
                     /map="60C8-60D2"
BASE COUNT    37500 a  33853 c  33980 g  38891 t   5368 others
ORIGIN


  Query Match             8.0%;  Score 64.6;  DB 33;  Length 149592;
  Best Local Similarity   55.6%;  Pred. No. 1.3e-06;
  Matches  124;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     530 ATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 589
           ||   |||||||||  |    |     || |   | | ||| | ||||| ||   ||| |
Db   25597 ATAGGAATGGCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTCG 25538

Qy     590 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAG 649
           | |  ||| |||| ||  | || || || |    | ||  |  | ||||| |||||  | |
Db   25537 GCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTTCTGCACCCTCG 25478

Qy     650 GGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGAC 709
           |          | ||    || |||||  || | || ||  |  |||||  |||| || |
Db   25477 GCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGATAGCCCTGCAGGTTGTGGGTC 25418

Qy     710 CTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAG 752
           |  | |||||       ||  | |||   ||||| ||| || |
Db   25417 CGGCGGATGGCTCCGCCTATGTGATACCCTACTACGGAGCCGG 25375


RESULT   2
AC007175/c
LOCUS       AC007175   188633 bp    DNA             INV       27-MAR-1999
DEFINITION  Drosophila melanogaster, chromosome 2R, region 57D3-57E2, BAC
            clones DS01261 and BACR48K03, complete sequence.
ACCESSION   AC007175 AC004434 AC003838 AC003531 AC003532 AC003533 AC006092
VERSION     AC007175.1  GI:4529971
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 188633)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2R, region 57D3-57E2
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 188633)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1999) Berkeley Drosophila Genome Project, MS-
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     On Mar 27, 1999 this sequence version replaced gi:4510440
            gi:4454570.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            Pl library location: 14-13.
FEATURES             Location/Qualifiers
     source          1. .188633
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="57D3-57E2"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
                     /clone="BACs DS01261 (D191) and BACR48K03 (D487)"
                     /note="DS01261 (D191) and R48K03 (D487) were sequenced as
                     a project.  R48K03 extends from BAC end at bp 1 to BAC end
                     at bp 132,637.  DS01261 extends from Pl end at bp 126,093
                     to a minimal overlap with its distal neighbor DS08012
                     (D224) at bp 188,633.  "
BASE COUNT    53352 a  41815 c  41597 g  51869 t
ORIGIN


  Query Match             8.0%;  Score 64.6;  DB 35;  Length 188633;
  Best Local Similarity   55.6%;  Pred. No. 1.3e-06;
  Matches  124;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     530 ATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 589
           ||   |||||||||  |    |     || |   | | ||| | ||||| ||   ||| |
Db 127193 ATAGGAATGGCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTCG 127134

Qy     590 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAG 649
           | |  ||| |||| ||  | || || || |    | ||  |  | ||||| |||||  | |
Db 127133 GCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTTCTGCACCCTCG 127074
```

```
Qy     650 GGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGAC 709
           |           | ||    || ||||| || | || ||  |  ||||| |||| || |
Db 127073 GCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGATAGCCCTGCAGGTTGTGGGTC 127014

Qy     710 CTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAG 752
           |  | |||||       ||  | |||   ||||| ||| || |
Db 127013 CGGCGGATGGCTCCGCCTATGTGATACCCTACTACGGAGCCGG 126971


RESULT   3
AC004434
ID   AC004434   standard; DNA; HTG; 83204 BP.
XX
AC   AC004434; AC003531; AC003532; AC003533; AC003838;
XX
SV   AC004434.4
XX
DT   18-MAR-1998 (Rel. 55, Created)
DT   25-MAR-1999 (Rel. 59, Last updated, Version 17)

DE   Drosophila melanogaster; Chromosome 2R; Region 57D11-57E2; P1 clone
DE   DS01261, WORKING DRAFT SEQUENCE, 10 unordered pieces.
XX
KW   HTG; HTGS_PHASE1.
XX
OS   Drosophila melanogaster (fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC   Drosophilidae; Drosophila.
XX
RN   [1]
RP   1-83204
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster";
RL   Unpublished.
XX
RN   [2]
RP   1-83204
RA   Celniker S.E., George R.A., Galle R., Svirskas R.R., Hoskins R.A.,
RA   Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Chavez C., Chew M.,
RA   Doyle C.M., Farfan D.E., Flanagan J., Houston K.A., Hummasti S.R.,
RA   Karra K., Kearney L., Kim S.H., Lee B., Lomotan M.A., Mak J., Mazda P.,
RA   Mok M.S., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S.,
RA   Pfeiffer B., Punch E., Snir E., Twomey B., Wan K.H., Whitelaw K.R., Yee A.,
RA   Zhang R., Zieran L.L., Kimmel B.E.;
RT   ;
RL   Submitted (17-MAR-1998) to the EMBL/GenBank/DDBJ databases.
RL   Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,
RL   Berkeley, CA 94720, USA
XX
CC   For further information about this sequence, including its location
CC   and relationship to other sequences, please visit our sequence
CC   archive Web site (http://www.fruitfly.org/sequence/) or send email
CC   to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
CC   the following cutoffs: length >= 400 bases, phrap computed error
CC   rate <= 1/10. P1 library location: 14-13.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 10 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *       1      738: contig of 738 bp in length
CC   *     739      818: gap of unknown length
CC   *     819     1918: contig of 1100 bp in length
CC   *    1919     1998: gap of unknown length
CC   *    1999     2644: contig of 646 bp in length
CC   *    2645     2724: gap of unknown length
CC   *    2725     3214: contig of 490 bp in length
CC   *    3215     3294: gap of unknown length
CC   *    3295     4442: contig of 1148 bp in length
CC   *    4443     4522: gap of unknown length
CC   *    4523     5169: contig of 647 bp in length
CC   *    5170     5249: gap of unknown length
CC   *    5250     6692: contig of 1443 bp in length
CC   *    6693     6772: gap of unknown length
CC   *    6773     9720: contig of 2948 bp in length
CC   *    9721     9800: gap of unknown length
CC   *    9801    13304: contig of 3504 bp in length
CC   *   13305    13384: gap of unknown length
CC   *   13385    83204: contig of 69820 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .83204
FT                   /chromosome="2"
FT                   /db_xref="taxon:7227"
FT                   /organism="Drosophila melanogaster"
FT                   /strain="y2; cn bw sp"
FT                   /clone="DS01261 (D191)"
FT                   /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
FT                   /map="57D11-57E2"
XX
SQ   Sequence 83204 BP; 23971 A; 17293 C; 17121 G; 24097 T; 722 other;


  Query Match             8.0%;  Score 64.6;  DB 46;  Length 83204;
  Best Local Similarity   55.6%;  Pred. No. 1.1e-06;
  Matches  124;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     530 ATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 589
           ||   |||||||||  |    |     || |   | | ||| | ||||| ||   ||| |
Db  82110 ATAGGAATGGCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTCG 82169

Qy     590 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAG 649
           | |  ||| |||| ||  | || || || |   | ||  |  | ||||| |||||  | |
Db  82170 GCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTTCTGCACCCTCG 82229

Qy     650 GGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGAC 709
           |          | ||    || |||||  || | || ||  |  ||||| |||| || |
Db  82230 GCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGATAGCCCTGCAGGTTGTGGGTC 82289

Qy     710 CTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAG 752
           |  | |||||       ||  | |||   ||||| ||| || |
Db  82290 CGGCGGATGGCTCCGCCTATGTGATACCCTACTACGGAGCCGG 82332


RESULT   4
AC006092
ID   AC006092   standard; DNA; HTG; 132637 BP.
XX
AC   AC006092;
XX
SV   AC006092.3
XX
DT   07-DEC-1998 (Rel. 57, Created)
DT   29-MAR-1999 (Rel. 59, Last updated, Version 10)
XX
DE   Drosophila melanogaster; Chromosome 2R; Region 57D3-57D12; BAC
DE   clone BACR48K03, WORKING DRAFT SEQUENCE, 1 unordered pieces.
XX
KW   HTG; HTGS_PHASE1.
XX
OS   Drosophila melanogaster (fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC   Drosophilidae; Drosophila.
```

```
XX
RN   [1]
RP   1-132637
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster";
RL   Unpublished.
XX
RN   [2]
RP   1-132637
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Punch E.,
RA   Sequeira A., Sethi H., Snir E., Svirskas R.R., Twomey B., Wan K.H.,
RA   Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;
RT   ;
RL   Submitted (04-DEC-1998) to the EMBL/GenBank/DDBJ databases.
RL   Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,
RL   Berkeley, CA 94720, USA
XX
CC   On Mar 20, 1999 this sequence version replaced gi:4417233.
CC   For further information about this sequence, including its location
CC   and relationship to other sequences, please visit our sequence
CC   archive Web site (http://www.fruitfly.org/sequence/) or send email
CC   to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
CC   the following cutoffs: length >= 400 bases, phrap computed error
CC   rate <= 1/10.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 1 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1   132637: contig of 132637 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .132637
FT                   /chromosome="2"
FT                   /db_xref="taxon:7227"
FT                   /organism="Drosophila melanogaster"
FT                   /strain="y2; cn bw sp"
FT                   /clone="BACR48K03 (D487) RPCI-98 48.K.3"
FT                   /clone_lib="RPCI-98 (Roswell Park Cancer Institute
FT                   Drosophila melanogaster BAC library, partial EcoRI in
FT                   pBACe3.6"
FT                   /map="57D3-57D12"
XX
SQ   Sequence 132637 BP; 35643 A; 30071 C; 30187 G; 36736 T; 0 other;


  Query Match             8.0%;  Score 64.6;  DB 46;  Length 132637;
  Best Local Similarity   55.6%;  Pred. No. 1.2e-06;
  Matches  124;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     530 ATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 589
           ||   |||||||||  |    |      || |   | | ||| | ||||| ||   ||| |
Db    5446 ATAGGAATGGCTATCACCTGGACACCACACTGCTACTCTCTGTCTTTCTGGGCATGTTCG 5505

Qy     590 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAG 649
           | |  ||| |||| ||  | || || || |   | ||  |  | ||||| |||||  | |
Db    5506 GCGTGGATAGATTCTATTTGGGCTATCCGGGCATCGGACTCCTCAAGTTCTGCACCCTCG 5565

Qy     650 GGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGAC 709
           |           | ||    || |||||  || | || ||  |  ||||| |||| || |
Db    5566 GCGGCATGTTCCTGGGCCAGCTGATTGACATCGTGCTGATAGCCCTGCAGGTTGTGGGTC 5625

Qy     710 CTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAG 752
           |  | |||||       ||  | |||   ||||| ||| || |
Db    5626 CGGCGGATGGCTCCGCCTATGTGATACCCTACTACGGAGCCGG 5668


RESULT   5
CELC41D11/c
LOCUS       CELC41D11   40912 bp    DNA             INV       27-FEB-1998
DEFINITION  Caenorhabditis elegans cosmid C41D11.
ACCESSION   AF003740
VERSION     AF003740.1  GI:2105490
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40912)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 40912)
  AUTHORS   Gattung,S and Maggi,L.
  TITLE     The sequence of C. elegans cosmid C41D11
  JOURNAL   Unpublished (1998)
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 1RQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone


                        NEIGHBORING COSMID INFORMATION:


            The 5' cosmid is C11D9, 3100 bp overlap.  Actual start of this
            cosmid is at base position 1 of CELC41D11; actual end is at 40912
            of CELC41D11

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
```

```
FEATURES             Location/Qualifiers
     source          1. .40912
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="C41D11"
     gene            complement(6009. .8933)
                     /gene="C41D11.3"
     CDS             complement(join(6009. .6173,6597. .6821,7005. .7699,
                     7830. .8157,8748. .8933))
                     /gene="C41D11.3"
                     /note="weak similarity to rat mitochondrial capsule
                     selenoprotein (GB:X87883) and integrin beta
                     oligodendroglia chain precursor (NID:g520566); coded for
                     by C. elegans cDNA yk2d8.3; coded for by C. elegans cDNA
                     yk130f5.3; coded for by C. elegans cDNA yk101b2.3; coded
                     for by C. elegans cDNA yk121e1.3; coded for by C. elegans
                     cDNA yk130f5.5; coded for by C. elegans cDNA yk101b2.5;
                     coded for by C. elegans cDNA yk121e1.5"
                     /codon_start=1
                     /protein_id="AAC48138.1"
                     /db_xref="GI:2105493"
                     /translation="MVPLVKKQMSIDEKIMEDDEDEGFGDSMKTIKRGQGSSLARSLS
                     LKLSPKTDQQKPIGADTEKSCSTPSLPTAPLLSRQGSTSSLRFKVSPSTFLKPIATPI
                     IDVSPTPRPDPEPLRIIVAAAAAAPVESTELSEPSTSKTPEQKAEPAMQKSALRKHAA
                     NFSVREEIFIPGIRRKKIRFSGLHVHYFDRRQGGSTVPTEGDISLDMHNKHHSHRYFS
                     LSTGKRPILDLLLYADDDLSDGEVIDPDSDEDRDYDEYASAKSIPRINQRHRIKLLKK
                     SGVKVEKNTDAIDSLRNTRVECGCSCENGVCLPETCQCAIDGIHCQVDGGEWPTQPCA
                     CFAETCTNPEGRTFYDPEAVHEFRQRTIMNWRASQKTGISGSPVVKKFADSDEEDDKA
                     KTWLLKNAIPIKLEESPTKYPVTPVYTRRSRSSLSDIRESTSETTGSITPSTSLSERI
                     KAMQHLDNEVCVEQEEKDEDDAYFMDNKEDTIIEADLTLQEVYVEQDMVAVTVESSSE
                     TTDIDANEMIEATAMSPVKTTLVV"
     gene            complement(18663. .20974)
                     /gene="C41D11.4"
     CDS             complement(join(18663. .18779,18829. .18944,19622. .19868,
                     20561. .20974))
                     /gene="C41D11.4"
                     /note="coded for by C. elegans cDNA yk21e1.5; coded for by
                     C. elegans cDNA yk2d8.5"
                     /codon_start=1
                     /protein_id="AAC48139.1"
                     /db_xref="GI:2105494"
                     /translation="MVSVAASLKRAGVPSLTRTKTSALPIKMEKKPMVPLRIEQDHQT
                     PIRKLAPESIPSLEDQVTPPMRRSSSGSMLLRQTPERIRRQEEDVETPGSPVKRYRRS
                     GSMPRLEAQSLDVPIRMEIGNSLTPVLKRKRTSSLTCLFSSDMPPILTPQRLRPRGIS
                     ETRTIDDDVFVPKSPTKKPITSPIKKRVRQDSENRSTSPLKKRYISSFSIRLEEEKEI
                     EKAVPTPTMSTPMRKFKSTIFNPRAVTERIAQTYGTGSRQLLKSTSSTVLPAPPALVA
                     RSTSGETYYSATEGGLKFFKY"
     gene            21167. .23273
                     /gene="C41D11.2"
     CDS             join(21167. .21266,21736. .21863,21909. .22031,22078. .22273,
                     22321. .22420,22470. .22941,23007. .23273)
                     /gene="C41D11.2"
                     /note="coded for by C. elegans cDNA yk83c1.3; coded for by
                     C. elegans cDNA yk83c1.5"
                     /codon_start=1
                     /protein_id="AAC48137.1"
                     /db_xref="GI:2105492"
                     /translation="MVSETNLQLFFNMHLVIILNLQMDSNKASLFKNWLSKMSTAVTI
                     TAPSVKHILLDSLVVMKIVKHVDSELHAGISEVSGDACAGVLTGLVFLEDSRLEITNC
                     FPTVRNEPVMDDDANAAQQYEEQKQHEMLDMLRKFRTMNIDYEIVGFYQSHQFGAGFS
                     HDLVESMFDYQAMGPENVVLIYDPIKTRQGQLSLRAWRLSTAALDLASKNDWRPELVK
                     AAGLTYQNMFEELPIIIKSSYLNNVLMSELSLAKSCSSDKYSTRHFDLGSKKSLEKSV
                     RAMMANVDELNKSIQSLTKYTIDKQRHDNMVFSLTQKRVRISNFEIKNILIVSLKFIF
                     RSVTAFKPKVLFFENHFPTYRGSTNYKTGCEGYIFGGIKQQENESRVARGDPTIPMDD
                     IKRIKAPQLQTRNGLLDELLASFDTNALADFSKTVTSENITKMFIAEAVAEEKVAGTK
                     DRTLSSVSSTR"
     gene            complement(23679. .27940)
                     /gene="C41D11.5"
     CDS             complement(join(23679. .24030,24286. .24440,24497. .24619,
                     25129. .25305,25352. .25478,25755. .26011,26081. .26330,
                     26721. .26791,27033. .27433,27495. .27600,27647. .27731,
                     27783. .27940))
                     /gene="C41D11.5"
                     /note="similar to the DNA/RNA non-specific endoneuclease
                     family; coded for by C. elegans cDNA CEESU20R"
                     /codon_start=1
                     /protein_id="AAC48141.1"
                     /db_xref="GI:2105496"
                     /translation="MMSLSSLNNENLNDEIIDFESKCKRFPVMFKRAVCIHSSLQFVD
                     LFMQNLMFSHGFSSTPIIARQEDETEVLSKTMLSFYKIRNEFREIFRSHHRGNVSEVA
                     IFIGTCPVRSTYSYRVPITICEAEDSTHKSCGDTCAELNDFERRKINRDLFMHSCSKE
                     QQSNETNKTNKNHRLFVFIKGSEDMVNEEIEEELVHEFPPNFFFLFFFFRDPPFGSKQ
                     SFHYSFIHEKCNCDGVPAPRNDGKWLRMVPYIMLHKILFLICLASFIPTIGSISGTKD
                     VKSKNCDGSAGLTCTFPGDCRIGDTVKVNCTSRKGCPNPVSRNNVEAVCRFCWQLLPG
                     DYDCEPATNCSTSSTKLLVTKCSAHSSVICMGQRNFYKRIPCNWSSGYSWTKTMILSV
                     VLGGFGADRFYLGLWKSAIGKLFSFGGLGVWTLVDVVLIAVGYIKPYDGSIRNMIGKV
                     AGTAAIAGISFLAGKYSNDDLPIFRNVQSATNVPMNQIQVSEPMTVKPASLNADAMGP
                     SRSAEIMKHGYPGFTNVRTYEDFVLSYDYKTRTAHWVCEHLTPERLKHAEGVDRKLCE
                     FKPDITFPQKFLSQNTDYKCSGFDRGHLAAAGNHRKSQLAVDQTFYLSNMSPQVGRGF
                     NRDKWNDLEMHCRRVAKKMINSYIITGPLYLPKLEGDGKKYIKYQVIGDNNVAVPTHF
                     FKVALFEVTPGKFELESYILPNAVIEDTVEISKFHVPLDAVERSAGLEIFARLDPKSI
                     VKENGAKKGGLLW"
     gene            complement(28522. .33822)
                     /gene="C41D11.7"
     CDS             complement(join(28522. .28843,29286. .29418,29463. .29705,
                     30317. .30489,30536. .30718,31647. .31861,31908. .32009,
                     32388. .32461,32961. .33352,33758. .33822))
                     /gene="C41D11.7"
                     /note="similar to mammalian DNA-binding proteins SMUBP-2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC48142.1"
                     /db_xref="GI:2105497"
                     /translation="MIEADKKFLELVKKWGVLSVGLNQKQNTNPVYQYTLEDNPESLT
                     LQDVLESHMTDPIELNSFDSMEKEAGIPFSNEKQRTAIKMALNENRKLVCIQGPPGTG
                     KTFTLTLLLCRLIQQKKQVVVLAPTREALANIRMMTKKTLKRMGIKVHEHALMDTNEY
                     RDVINKSDRALMAAEEVRDLRKAFDNGEITENVLDEMRQSIINRVRNEVGAEVIGNVR
                     VAFATIGASFVDFVMKHKKFDPCLCIIDEAAQVMEAQTWPAVYKMKRIVMAGDPKQLP
                     ALVFTDEAKAFGLQNSVMDRILEKKNNFSWIMLENQYRSNAKIATWSNTCFYHNQLKT
                     DVKCHEYSLHTILNPQPKKFRNLFDPLVLIDTSLERDVEKRLETYEHAVFDTNSINKT
                     KQGFSYANLAEAKIAIGHYQRLLKYGVQPSDIAIITPYKGQTSLVTKLMEEFGAETGY
                     TDFVQTTIGTVDSVQGKEYEVVIFTMVRSNPRKTMGFVSELRRLNVVITRAKRHFMFI
                     GNGYLLAESKKDEIRKLYDCFKNAKRRFHPNNAFGEDGNVTDYVSNNFGQDLESFMNY
                     SNDEEMINWCKRFNENGPDYRAKKALAWEKREETRLMKTISKVKYEYKRGSTEQLASE
                     LDQLDCSSN"
     gene            34408. .35688
                     /gene="C41D11.1"
     CDS             join(34408. .34592,34661. .34754,34805. .34917,34967. .35153,
                     35210. .35348,35399. .35505,35551. .35688)
                     /gene="C41D11.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC48136.1"
                     /db_xref="GI:2105491"
                     /translation="MTDKENFYSGFFLGRLPGSFDSRFLIKVACFQPGKEERNVFELT
                     TDTKREEDYGEMVHFWVDKGPLQNKSFVRFEENYETFFIHENKMFDRIVHTDKDFYVG
                     VNDLPDSDIYIHPDLGKMIDTRSIVTPGKKYNFITFEVKQNENNSFDIEIVGGEVQGK
                     EKNSEEPSAKEKIIQKMEQVHEVMKVEYEYNKNEETANWDVIGHLKLIKFYDKDLTSG
                     LRPVLRLMDNRHVALFKKQRETMPIKMIKVIEVDGIWKDDVNFVPQEGTLLPTDHDEK
                     TVVLLGNDYRHKFERLFKREDVFRLESNNTLTHLKFWQYGLEKI"
     gene            complement(37163. .39721)
                     /gene="C41D11.6"
     CDS             complement(37163. .39721)
                     /gene="C41D11.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC48140.1"

  Query Match          6.1%;  Score 49.6;  DB 35;  Length 40912;
  Best Local Similarity  49.2%;  Pred. No. 0.01;
  Matches  130;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy    497 AAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGG 556
          ||  |  ||| | |||||  ||  | ||         |  ||| ||    |  | |
```

```
Db  25983 AACGTAATTTCTACAAGCGAATTCCATGTAACTGGTCATCTGGATACAGTTGGACAAAAA 25924

Qy    557 CAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACC 616
          || |     | ||  || | || ||  | || ||||| ||||| || ||| | |||  |
Db  25923 CAATGATTCTCTCAGTTGTGCTCGGCGGTTTTGGAGCTGATCGTTTCTACTTGGGACTCT 25864

Qy    617 CTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTG 676
          |  |   |   |||  ||| || | | ||   |  |||  |  |||  ||  | |
Db  25863 GGAAATCTGCAATTGGAAAACTTTTCAGTTTCGGTGGTCTTGGTGTGTGGACTCTTGTGG 25804

Qy    677 ATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAG 736
          |  |  |||| ||  || |        ||  |||||  ||||||   |  || |||  |
Db  25803 ACGTTGTTCTGATCGCAGTTGGATACATTAAACCTTACGATGGATCCATGTATATTTGAT 25744

Qy    737 ATTACTATGGAACCAGACTTACAA 760
            ||  | |  |       || |||
Db  25743 TTTTTTTTAATTTTCTTTTCCAA 25720


RESULT   6
AC016672
LOCUS       AC016672   205157 bp    DNA             HTG       04-DEC-1999
DEFINITION  Homo sapiens clone RP11-14C10, WORKING DRAFT SEQUENCE, 1 unordered
            pieces.
ACCESSION   AC016672
VERSION     AC016672.1  GI:6524410
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 205157)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205157)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT      Center project name: H_NH0014C10.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   205157: contig of 205157 bp in length.
FEATURES            Location/Qualifiers
     source         1. .205157
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="RP11-14C10"
BASE COUNT    53433 a  46205 c  47212 g  58163 t    144 others
ORIGIN


  Query Match             5.8%;  Score 47.2;  DB 44;  Length 205157;
  Best Local Similarity   56.4%;  Pred. No. 0.061;
  Matches   88;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy    577 CTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAG 636
          || ||  |||| |||||||| || || ||||| ||  |          ||  |    |||
Db  15774 CTCGGTGGGTTTGGAGCAGACCGTTTCTACCTGGGCCAGTGGCGGGAAGGCCTCGGCAAG 15833

Qy    637 TTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATG 696
          | | ||   | ||       |||||  |||  || || ||  || | || |||  | |
Db  15834 CTCTTCAGCTTCGGTGGCCTGGGAATATGGACGCTGATAGACGTCCTGCTCATTGGAGTT 15893

Qy    697 CAGATTGTTGGACCTTCAGATGGAAGTAGTTACATT 732
                ||||||||  |||||||   |   ||||||
Db  15894 GGCTATGTTGGACCAGCAGATGGCTCTTTGTACATT 15929


RESULT   7
DMBR30C13
LOCUS       DMBR30C13  117674 bp    DNA             HTG       19-OCT-1999
DEFINITION  Drosophila melanogaster chromosome X clone BACR30C13 map 8D strain
            y; cn bw sp, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION   AL122025
VERSION     AL122025.1  GI:6090476
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 117674)
  AUTHORS   Madueno,E., de Pablos,B. and Modolell,J.
  TITLE     Sequencing the X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    Centro de Biologia Molecular Severo Ochoa, Consejo Superior de,
            Investigaciones Cientificas and Universidad Autonoma de Madrid,
            28049 Madrid, Spain.
REFERENCE   2  (bases 1 to 117674)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1999) European Drosophila Genome Sequencing
            Consortium
COMMENT     This is a 'working draft' sequence.  It currently
            consists of 153 contigs. The true order of the pieces is not known
            and their order in this sequence record is arbitrary. Gaps between
            the contigs are represented as runs of N, but the exact sizes of
            the gaps are unknown. This record will be updated with the finished
            sequence. 1       513: contig of 513 in length
            514       613: gap of unknown length
            614      1380: contig of 767 in length
            1381      1480: gap of unknown length
            1481      1995: contig of 515 in length
            1996      2095: gap of unknown length
            2096      2862: contig of 767 in length
            2863      2962: gap of unknown length
            2963      3754: contig of 792 in length
            3755      3854: gap of unknown length
            3855      4657: contig of 803 in length
            4658      4757: gap of unknown length
            4758      5330: contig of 573 in length
            5331      5430: gap of unknown length
            5431      5932: contig of 502 in length
            5933      6032: gap of unknown length
            6033      6474: contig of 442 in length
            6475      6574: gap of unknown length
            6575      7188: contig of 614 in length
            7189      7288: gap of unknown length
            7289      7513: contig of 225 in length
            7514      7613: gap of unknown length
            7614      8268: contig of 655 in length
            8269      8368: gap of unknown length
            8369      9346: contig of 978 in length
            9347      9446: gap of unknown length
            9447     10066: contig of 620 in length
            10067     10166: gap of unknown length
            10167     10814: contig of 648 in length
            10815     10914: gap of unknown length
            10915     11203: contig of 289 in length
            11204     11303: gap of unknown length
            11304     11941: contig of 638 in length
            11942     12041: gap of unknown length
            12042     12665: contig of 624 in length
            12666     12765: gap of unknown length
            12766     13410: contig of 645 in length
            13411     13510: gap of unknown length
            13511     13996: contig of 486 in length
```

13997 14096: gap of unknown length
14097 14677: contig of 581 in length
14678 14777: gap of unknown length
14778 15318: contig of 541 in length
15319 15418: gap of unknown length
15419 16047: contig of 629 in length
16048 16147: gap of unknown length
16148 16788: contig of 641 in length
16789 16888: gap of unknown length
16889 17502: contig of 614 in length
17503 17602: gap of unknown length
17603 18281: contig of 679 in length
18282 18381: gap of unknown length
18382 18819: contig of 438 in length
18820 18919: gap of unknown length
18920 19754: contig of 835 in length
19755 19854: gap of unknown length
19855 20699: contig of 845 in length
20700 20799: gap of unknown length
20800 21183: contig of 384 in length
21184 21283: gap of unknown length
21284 22021: contig of 738 in length
22022 22121: gap of unknown length
22122 23187: contig of 1066 in length
23188 23287: gap of unknown length
23288 23945: contig of 658 in length
23946 24045: gap of unknown length
24046 24616: contig of 571 in length
24617 24716: gap of unknown length
24717 25676: contig of 960 in length
25677 25776: gap of unknown length
25777 26190: contig of 414 in length
26191 26290: gap of unknown length
26291 27292: contig of 1002 in length
27293 27392: gap of unknown length
27393 28107: contig of 715 in length
28108 28207: gap of unknown length
28208 28729: contig of 522 in length
28730 28829: gap of unknown length
28830 29976: contig of 1147 in length
29977 30076: gap of unknown length
30077 30928: contig of 852 in length
30929 31028: gap of unknown length
31029 31903: contig of 875 in length
31904 32003: gap of unknown length
32004 32418: contig of 415 in length
32419 32518: gap of unknown length
32519 33132: contig of 614 in length
33133 33232: gap of unknown length
33233 33823: contig of 591 in length
33824 33923: gap of unknown length
33924 34827: contig of 904 in length
34828 34927: gap of unknown length
34928 35934: contig of 1007 in length
35935 36034: gap of unknown length
36035 36962: contig of 928 in length
36963 37062: gap of unknown length
37063 37679: contig of 617 in length
37680 37779: gap of unknown length
37780 38484: contig of 705 in length
38485 38584: gap of unknown length
38585 39541: contig of 957 in length
39542 39641: gap of unknown length
39642 40517: contig of 876 in length
40518 40617: gap of unknown length
40618 41307: contig of 690 in length
41308 41407: gap of unknown length
41408 42011: contig of 604 in length
42012 42111: gap of unknown length
42112 42828: contig of 717 in length
42829 42928: gap of unknown length
42929 43593: contig of 665 in length
43594 43693: gap of unknown length
43694 44391: contig of 698 in length
44392 44491: gap of unknown length
44492 45147: contig of 656 in length
45148 45247: gap of unknown length
45248 46798: contig of 1551 in length
46799 46898: gap of unknown length
46899 47830: contig of 932 in length
47831 47930: gap of unknown length
47931 48883: contig of 953 in length
48884 48983: gap of unknown length
48984 50341: contig of 1358 in length
50342 50441: gap of unknown length
50442 51797: contig of 1356 in length
51798 51897: gap of unknown length
51898 52237: contig of 340 in length
52238 52337: gap of unknown length
52338 53710: contig of 1373 in length
53711 53810: gap of unknown length
53811 54430: contig of 620 in length
54431 54530: gap of unknown length
54531 55595: contig of 1065 in length
55596 55695: gap of unknown length
55696 56555: contig of 860 in length
56556 56655: gap of unknown length
56656 57909: contig of 1254 in length
57910 58009: gap of unknown length
58010 59436: contig of 1427 in length
59437 59536: gap of unknown length
59537 60724: contig of 1188 in length
60725 60824: gap of unknown length
60825 62193: contig of 1369 in length
62194 62293: gap of unknown length
62294 63069: contig of 776 in length
63070 63169: gap of unknown length
63170 64233: contig of 1064 in length
64234 64333: gap of unknown length
64334 64782: contig of 449 in length
64783 64882: gap of unknown length
64883 66244: contig of 1362 in length
66245 66344: gap of unknown length
66345 67507: contig of 1163 in length
67508 67607: gap of unknown length
67608 68877: contig of 1270 in length
68878 68977: gap of unknown length
68978 70692: contig of 1715 in length
70693 70792: gap of unknown length
70793 71673: contig of 881 in length
71674 71773: gap of unknown length
71774 72809: contig of 1036 in length
72810 72909: gap of unknown length
72910 74298: contig of 1389 in length
74299 74398: gap of unknown length
74399 76722: contig of 2324 in length
76723 76822: gap of unknown length
76823 79379: contig of 2557 in length
79380 79479: gap of unknown length
79480 79862: contig of 383 in length
79863 79962: gap of unknown length
79963 80475: contig of 513 in length
80476 80575: gap of unknown length

Query Match 5.7%; Score 46.2; DB 32; Length 117674;
Best Local Similarity 56.1%; Pred. No. 0.1;
Matches 87; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy    577 CTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAG 636
          || ||  | || ||||| |||||||| ||| | ||  ||          ||  |   ||||
Db   4279 CTAGGGGGCTTTGGAGCCGATCGATTCTACTTGGGCCACTGGCAGGAAGGGATTGGAAAG 4338

Qy    637 TTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATG 696
            | | ||   | ||        ||  |  ||| | | || ||| || | || || || |||
Db   4339 CTGTTCAGCTTTGGCGGCCTCGGCGTCTGGACCATCATCGATGTCCTGCTCATCTCGATG 4398
```

```
Qy     697 CAGATTGTTGGACCTTCAGATGGAAGTAGTTACAT 731
           ||   | | || ||  | |||||      ||||||
Db    4399 CATTATTTGGGGCCAGCGGATGGCTCACTTTACAT 4433


RESULT   8
AF181623
LOCUS       AF181623     1440 bp    mRNA            INV       16-SEP-1999
DEFINITION  Drosophila melanogaster BcDNA.GH02340 (BcDNA.GH02340) mRNA,
            complete cds.
ACCESSION   AF181623
VERSION     AF181623.1  GI:5901801
KEYWORDS    FLI_CDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1440)
  AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
            Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
            Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
            Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
            Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
  TITLE     Full Length Drosophila melanogaster cDNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1440)
  AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
            Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
            Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
            Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
            Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Berkeley Drosophila Genome Project,
            University of California Berkeley, 539 Life Sciences Addition
            #3200, Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            University of California Berkeley
            Berkeley, CA 94720
            For further information about this sequence,including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .1440
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH02974"
     gene            1. .1440
                     /gene="BcDNA.GH02974"
                     /note="motif=(desc:'C_TYPE_LECTIN_1', acc:'PS00615',
                     count:'1', method:'PPSEARCH', dbase:'PROSITE')"
     CDS             276. .1130
                     /gene="BcDNA.GH02974"
                     /codon_start=1
                     /product="BcDNA.GH02974"
                     /protein_id="AAD55409.1"
                     /db_xref="GI:5901802"
                     /db_xref="FLYBASE:FBcd0002974"
                     /translation="MRLQRQCIVVNMRSAIVLIMIFVLTGIRNSETASGGNQMDLSDS
                     KGDHKDNSNASNGNGNANDNEVYVPPLVSSMVAKSGGGAGGLLDNITAYSSSSSSSSS
                     NGNNNMLCPYDKETPCDRLQFPCIRCNYNHGCIYGRDLNVTCEVINNVQCLGERSFQR
                     QMNCRYCYQTEMWQQSCGQRSSCNSATDKLFRTNCTVHHDVLCLGNRSFTRNLRCNWT
                     QGYRWSTALLISLTLGGFGADRFYLGHWQEGIGKLFSFGGLGVWTIIDVLLISMHYLG
                     PADGSLYI"
BASE COUNT      422 a    325 c    329 g    364 t
ORIGIN
```

```
  Query Match             5.7%;  Score 46.2;  DB 35;  Length 1440;
  Best Local Similarity   56.1%;  Pred. No. 0.046;
  Matches   87;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy     577 CTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAG 636
           || ||  | || ||||| |||||||| ||| | ||  ||          ||  |   ||||
Db     972 CTAGGGGGCTTTGGAGCCGATCGATTCTACTTGGGCCACTGGCAGGAAGGGATTGGAAAG 1031

Qy     637 TTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATG 696
           .| | ||   | ||       ||  |  ||| | | || ||| || | || || || |||
Db    1032 CTGTTCAGCTTTGGCGGCCTCGGCGTCTGGACCATCATCGATGTCCTGCTCATCTCGATG 1091

Qy     697 CAGATTGTTGGACCTTCAGATGGAAGTAGTTACAT 731
           ||   | | || ||  | |||||      ||||||
Db    1092 CATTATTTGGGGCCAGCGGATGGCTCACTTTACAT 1126


RESULT   9
CELC02F5
LOCUS       CELC02F5    22333 bp    DNA             INV       22-OCT-1993
DEFINITION  C. elegans cosmid C02F5.
ACCESSION   L14745 L18807
VERSION     L14745.1  GI:289607
KEYWORDS    GRR1 gene; GTP-binding protein; glucose repression; proteasome
            component C5.
SOURCE      Caenorhabditis elegans (strain Bristol N2) hermaphrodite mixed
            whole animal DNA.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 22333)
  AUTHORS   Sulston,J., Du,Z., Thomas,K., Wilson,R., Hillier,L., Staden,R.,
            Halloran,N., Green,P., Thierry-Mieg,J., Qiu,L., Dear,S.,
            Coulson,A., Craxton,M., Durbin,R., Berks,M., Metzstein,M.,
            Hawkins,T., Ainscough,R. and Waterston,R.
  TITLE     The C.elegans genome sequencing project: A beginning
  JOURNAL   Nature 356, 37-41 (1992)
  MEDLINE   92168156
REFERENCE   2  (bases 1 to 22333)
  AUTHORS   Anderson,K.
  TITLE     Sequence of the C. elegans cosmid C02F5
  JOURNAL   Unpublished (1993)
COMMENT     Submitted by:
            Nematode Sequencing Project
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
            MRC Laboratory of Molecular Biology,
            Cambridge CB2 2QH, England
            e-mail: rw@nematode.wustl.edu and jes@cele.mrc-lmba.cam.ac.uk NOTE:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder (P. Green and L. Hillier, in preparation)
            Neighboring cosmid information:
            Bases 1 to 800 of this entry correspond to bases 26944 to 27743 of
            GenBank entry CELC30A5 (L10990).  Bases 22034 to 22333 of this
            entry
            correspond to bases 1 to 300 of GenBank CELF09G8 (L11247).  The
            sequence
            of C. elegans cosmid C02F5 begins at position 18944 of GenBank
            CELC30A5
            (L10990), continues through this entry and ends at position 15818
            of
            GenBank CELF09G8 (L11247).  The sequence of C. elegans cosmid F09G8
            begins at position 22034 of this entry.
FEATURES             Location/Qualifiers
     source          1. .22333
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /dev_stage="mixed"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /map="3"
```

```
CDS             join(8. .124,175. .257,305. .694,742. .943,993. .1493,
                2247. .2423)
                /standard_name="C02F5.7 protein"
                /note="homology with glucose induced repressor, GRR1;
                putative"
                /codon_start=1
                /protein_id="AAA27922.1"
                /db_xref="GI:289614"
                /translation="METAERISAAASAASSRRAKRLAQQAHKTHPVIQAKQNQMYLIT
                TLSPAQVDNSLINRVLPKEVLLKVFSFLDTKALCRSAQVCRSWSILALDGSNWQRVDL
                FTFQRDVKTAVVENLARRCGGFLKELSLKGCENVHDSALRTFTSRCPNLEHLSLYRCK
                RVTDASCENLGRYCHKLNYLNLENCSSITDRAMKYIGDGCPNLSYLNISWCDAIQDRG
                VQILSNCKSLDTLILRGCEGLTENVFGSVEAHMGAIKKLNLLQCFQLTDITVQNIAN
                GATALEYLCMSNCNQISDRSLVSLGQHSHNLKVLELSGCTLLGDNGFIPLARGCRQLE
                RLDMEDCSLISDHTINSLANNCTALRELSLSHCELITDESIQNLASKHRETLNVLELD
                NCPQLTDSTLSHLRHCKALKRIDLYDCQNVSKEAIVRFQVYVDAVLSYDCIYVPCSTR
                STENWYKLTKLTKSDINRAPCAHRVEAQLWYSIKNVLKF"
exon            175. .257
                /note="putative"
                /number=9
exon            305. .694
                /note="putative"
                /number=10
exon            742. .943
                /note="putative"
                /number=11
exon            993. .1493
                /note="putative"
                /number=12
misc_feature    2786
                /note="possible trans-spliced leader site for C02F5.6"
CDS             join(2793. .2834,3022. .3473,3528. .3686,3737. .4091,
                4139. .4252,4326. .4463)
                /standard_name="C02F5.6 protein"
                /note="putative"
                /codon_start=1
                /protein_id="AAA27921.1"
                /db_xref="GI:289613"
                /translation="MAHTSDGWGAPYDNQTYVEAYEQLEIALLEPLDRILETANVEEF
                RPKFKNHQDPNDRKNKKNNDEEWRDSIYNIATDESDTDDHEQRKNFFQPPLQVQRNSF
                VKNTLMEFKRSSQIDISRLAVMGCGEMSLEKGICEYLGSFGTINVLSVDIDEPSLSIG
                QQLLGKAESIFLRKHLERNAEILAVETGLPVLMRSYVGDILEPDHRFADVDAIVSMEV
                VEHIPLPNAKKFVENVLGTLMPRIFIFSTPNHEYNAVFGMEPGEFRHGDHKFEMNRKE
                FSNWLEELSIRFPHYQIDPPHYIGMTRGYENLSGASQAAVCRLQVDLNTTLPQEVTPY
                EMVGHLPCRLGSRLIAYNLVKEAFLDWLEKIELQEHEPRTDGYSPYWIFNVQNILHHL
                KAPVSFALTIDEKVAMQVSSVIGTVEF"
exon            3022. .3473
                /note="putative"
                /number=13
exon            3528. .3686
                /note="putative"
                /number=14
exon            3737. .4091
                /note="putative"
                /number=15
exon            4139. .4252
                /note="putative"
                /number=16
CDS             join(6347. .6436,6481. .6795,6876. .6992)
                /standard_name="C02F5.5 protein"
                /note="putative"
                /codon_start=1
                /protein_id="AAA27920.1"
                /db_xref="GI:289612"
                /translation="MSYMEQIPPYEYYKYVFGTRCAVLACSSFELMMVLLGSLGDSNV
                LAKLFYLVFLGASAAVSGFNLQQNIDGRDEIKKITGSGDNETRGKAAALVGIPALAAV
                FVFLCVSGHAFFSLFVLVHVLASIGQLGIEAYEVRNGFCIKAPENKKAVADSYDQNYQ
                ILNMVQDSVFDKK"
exon            6481. .6795
                /note="putative"
                /number=17
CDS             complement(join(8274. .8477,8518. .8583,8631. .8897,
                9041. .9178,9473. .9532))
                /standard_name="C02F5.8 protein"
                /note="putative"
                /codon_start=1
                /protein_id="AAA27923.1"
                /db_xref="GI:289615"
                /translation="MATWKFIIRSVLFFLDLAMLLAALALIAVGFWMGYDSSFDTDLK
                NVIYKYDDPKSLADAKFNIRVWLIVVFWSIIGLSLGAVVTAVLGMISSVWPKRKGFMI
                TYLVLIIVLVSLEIGCGVAVLVRRNSLHDNTNSLIDAMYTTNSVNDLKIIQDKYNCCG
                IENSLFNVMYCGPMSQKPHCDVAVFDSVDNTMMISGIILLVILILQTIAIILPVPILI
                SRKKTYKYSYEPRVTQLADITEDTRF"
exon            8518. .8583
                /note="complement; putative"
                /number=6
exon            8631. .8897
                /note="complement; putative"
                /number=7
exon            9041. .9178
                /note="complement; putative"
                /number=8
misc_feature    9558
                /note="possible trans-spliced leader site for C02F5.8"
CDS             complement(join(11079. .11495,11809. .12165,12215. .12369,
                13369. .13474,13523. .13916,13962. .14227))
                /standard_name="C02F5.9 protein"
                /note="homology with proteasome component C5; coded for by
                C. elegans cDNAs GenBank:M80127 and T01498; putative"
                /codon_start=1
                /protein_id="AAA27924.1"
                /db_xref="GI:289616"
                /translation="MTSFTGITAVANATNEMAMFKQAMKEVAAHPEWMSSRQIERQRW
                NPYSMEGGSTCAISGENFAIVASDTRMTQNDINILTRDAEKIQILNDNIILTTSGFYG
                DVLQLKKVLQSRLHKYRFDYRSDMSVDLCAELLSRNLYYRRFFPYYTGAILAGIDEHG
                KGAVFSYDPIGCIERLGYSASGAAEPMIIPFLDCQIGHVTLSEGYERPELTLDRAISL
                MKDSFRGAAEREISTGDKIHLVIAEAGKPVVVKFLPLQSISAQSQNVLKKPFKSISAE
                LGFAQPLNLVDTLGLKPRDSDAMSRGKRPASVLKGLFDDWDPTANLDIPGTSSDIPSD
                PSSALKVPKKEVLDESEEILDQTSGSSSFSLNDSEQANEQERNIIEDLLRQQMFTESK
                IKNEEASLLKGLKKGLGEELESSHANYVITCNFPGCGLKYNWRVKYGKLRLLDHALTH
                SNRKIPCKLCGFECTNVRRMRSHYAKAHPNERVEGYGMKALVSGDSSRIGDGVDGDID
                QQVSDDELKELWNNCFSESIHLVGQASGFVEGEKYRRMTKRRKLDREAMNSMNFMF"
exon            11809. .12165
                /note="complement; putative"
                /number=2
exon            12215. .12369
                /note="complement; putative"
                /number=3
exon            13369. .13474
                /note="complement; putative"
                /number=4
exon            13523. .13916
                /note="complement; putative"
                /number=5
CDS             join(14580. .14717,14761. .14894,14937. .15120,15330. .15412,
                15465. .15612,15661. .15759)
                /standard_name="C02F5.4 protein"
                /note="coded for by C. elegans cDNA GenBank:CEL16D4
                (M89109); putative"
                /codon_start=1
                /protein_id="AAA27919.1"
                /db_xref="GI:289611"
                /translation="MADFTEQTLRQKLTNLSNHPSSIQTTSAWLLQNHSNRELIIRVW
                LKTVKKETHGSKIVNLLYVANDVSQNARKTCPQFKDEFFPAIESSFRHAIELKNAKEV
                EHAIGKLINVWKDRQIFTPSQCKRLHEVHQQVKLSGSFPTPAVANKEHGKSVRSSQEA
                QSLLVQNEEALPMLEEYVKRLKNETNERETLESNLNMLIENVRMSIEHHEKLCREVKR
                REDRIKADLLEVEKTFESLPDLAAEMPNAPLPTLEALFEKRKK"
exon            14761. .14894
                /note="putative"
                /number=18
exon            14937. .15120
                /note="putative"
                /number=19
exon            15330. .15412
                /note="putative"
                /number=20
```

```
     exon            15465. .15612

  Query Match              5.3%;  Score 43.2;  DB 34;  Length 22333;
  Best Local Similarity   54.4%;  Pred. No. 0.48;
  Matches   87;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
           ||  ||||||| ||   | | |  |  || || |||| ||| |||||  ||||  ||| |
Db   18612 TCCATTTTTCTCGGTGTGGTTGCTGTCGACCGTTTTTGCCTGGGATATTCTGCAATGGCT 18671

Qy     628 TTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 687
           |   |||  |    ||| | ||    | ||||||  || |  |  | ||| |  |  |
Db   18672 GTCGGAAAACTAATGACTTTGGGTGGATTTGGAATCTGGTGGATTGTCGATATATTCTTA 18731

Qy     688 ATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTT 727
           |   | |     |  | |||||  | ||||     ||||
Db   18732 CTGGTACTTGGAGTACTCGGACCCGCGGATGACTCGAGTT 18771


RESULT  10
Y15035/c
LOCUS       CVY15035    49649 bp    DNA             VRL       29-APR-1998
DEFINITION  Cowpox virus strain GRI-90 DNA (49 kb fragment).
ACCESSION   Y15035
VERSION     Y15035.1  GI:3097015
KEYWORDS    .
SOURCE      Cowpox virus.
  ORGANISM  Cowpox virus
            Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
            Orthopoxvirus.
REFERENCE   1  (bases 1 to 49649)
  AUTHORS   Shchelkunov,S.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1997) S.N. Shchelkunov, Institute of Molecular
            Biology, 633159 Novosibirsk region, Koltsovo, SRCVB 'Vector', USSR
REFERENCE   2  (bases 1 to 49649)
  AUTHORS   Shchelkunov,S.N., Safronov,P.F., Totmenin,A.V., Petrov,N.A.,
            Ryazankina,O.I., Gutorov,V.V. and Kotwal,G.J.
  TITLE     The genomic sequence analysis of the left and right
            species-specific terminal region of a cowpox virus strain reveals
            unique sequences and a cluster of intact ORFs for immunomodulatory
            and host range proteins
  JOURNAL   Virology 243 (2), 432-460 (1998)
  MEDLINE   98229462
COMMENT     Overlaps with sequence published in FEBS Lett., 327, 321-324, 1993.
FEATURES             Location/Qualifiers
     source          1. .49649
                     /organism="Cowpox virus"
                     /strain="GRI-90"
                     /db_xref="taxon:10243"
     gene            228. .1232
                     /gene="A51R"
     CDS             228. .1232
                     /gene="A51R"
                     /codon_start=1
                     /protein_id="CAA75271.1"
                     /db_xref="GI:3097016"
                     /db_xref="SPTREMBL:O72733"
                     /translation="MDGVIVYCLNALVKHGEEINHIKNDFMIKPCCERVCKKVKNVHI
                     DGQSKNNTVIADLPYLDNAVSDVCKSIYKKNVSRISRFANLIKIDDDDKTPTGVYNYF
                     KPKDAIPVIISIGKDKDVCELLIASDKACACIELNSYKVAILPMDVSFFTKGNASLII
                     LLFDFSIDAAPLLRSVTDNNVIISRHKRLHDELPSSNWFKFYISIKSDYCSILYMVVD
                     GSVMYAIADNRTHAIISKNILDNTTINDECRCCYFEPQIKILDRDEMLNGSSCDMNRH
                     CIMMNLPDVGEFGSSMLGKYEPDMIKIALSVAGNLIRNRDYIPGRRGYSYYVYGIASR
                     "
     gene            1302. .1874
                     /gene="A52R"
     CDS             1302. .1874
                     /gene="A52R"
                     /codon_start=1
                     /protein_id="CAA75272.1"
                     /db_xref="GI:3097017"
                     /db_xref="SPTREMBL:O72734"
                     /translation="MDIKIDISISGDKFTVTTRRENEERKKYLPLQKEKTTDVIKPDY
                     LEYDDLLDRDEMSTILEEYFMYRGLLGLRIKYGRLFNEIRKFDNDAEEQFGTIEELKQ
                     KLRLNSEEGADNFIDYIKVQKQDIVKLTVYDCISMIGLCACVVDVWRKEKLFSRWKYC
                     LRAIKLFIDDHMLDKIKSILQNRLVYVEMS"
     gene            2198. .2758
                     /gene="A53R"
     CDS             2198. .2758
                     /gene="A53R"
                     /codon_start=1
                     /protein_id="CAA75273.1"
                     /db_xref="GI:3097018"
                     /db_xref="SPTREMBL:O72735"
                     /translation="MDIKSLLAVCTILYITTLVTADIPTPLPPHAPVNGSCDEGEYLD
                     KRHNQCCNQCPPGEFAKVRCSGSDNTKCERCPPHTYTAIPNYSNGCHQCRKCPTGSFD
                     KVKCTGTQNSKCSCLPGWYCATDSSQTEDCRDCVPKSRCPCGYFGGIDEQGNPICKSC
                     CVGEYCDYLRNYRLDPFPPCKLSKCN"
     gene            3015. .4709
                     /gene="A54R"
     CDS             3015. .4709
                     /gene="A54R"
                     /codon_start=1
                     /protein_id="CAA75274.1"
                     /db_xref="GI:3097019"
                     /db_xref="SPTREMBL:O72736"
                     /translation="MNNSSELIAVINGFRNSGRFCDINIVINDERINAHRLILSGASE
                     YFSILFSNNFIDSNEYEVNLSHLDYQSVNDLIDYIYGIPLSLTNDNVKYILSTADFLQ
                     IGSAITECEKYILKNLCSRNCIDFYIYADKYNNKIESASFNTILRNILRLINDENFK
                     YLTEESMIKILSDDMLNIKNEDFAPLILIKWLESTQQSCTVELLRCLRISLLSPQVIK
                     SLYSHQLVSSIYECITFLNNIAFLDESFPRYHSIELISIGISNSHDKISINCYNHKKN
                     TWEMISSRRYRCSFAVAVLDNIIYMMGGYDQSPYRSSKVIAYNTCTNSWIYDIPELKY
                     PRSNCGGLADDEYIYCIGGIRDQDSSLTSSIDRWKPSKPYWQKYAKMREPKCDMGVAM
                     LNGLIYVIGGIVKGDTCTDALESLSEDGWMKHQRLPIKMSNMSTIVHDGKIYISGGYN
                     NSSVVNVISNLVLSYNPIYDEWTKLSSLNIPRINPALWSAHNKLYVGGGISDDVRTNT
                     SETYDKEKDCWTLDNGHVLPRNYIMYKCEPIKHKYPLEKTQYTNDFLKYLESFIGS"
     gene            4759. .5703
                     /gene="A55R"
     CDS             4759. .5703
                     /gene="A55R"
                     /codon_start=1
                     /protein_id="CAA75275.1"
                     /db_xref="GI:3097020"
                     /db_xref="SPTREMBL:O72737"
                     /translation="MARLPILLLLISLVYSTPSPQTSKKIGDDATLSCNRNNTTDYVV
                     MSAWYKEPNSIILLAAKSDVLYFDNYTKDKISYDSPYDDLVTTITIKSLTARDAGTYV
                     CAFFMTSTTNDTDKVDYEEYSTELIVNTDSESTIDIILSGSTHSPETSSEKPDYIDNS
                     NCSSVFEIATPEPITDNEEDHTDVTYTSENINTVSTTSRESTTDETPEPITDKEEDHT
                     VTDTVSYTTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSISNYKTKDF
                     VEIFGITALIILSAVAIFCITYYICNKRSRKYKTENKV"
     gene            5721. .6314
                     /gene="A56R"
     CDS             5721. .6314
                     /gene="A56R"
                     /codon_start=1
                     /protein_id="CAA75276.1"
                     /db_xref="GI:3097021"
                     /db_xref="SPTREMBL:O72738"
                     /translation="MSGIVKSIILSGPSGSGKTAIAKRLLKDYGNIFGFVVSHTTRFP
                     RPMEREGVDYHYVNREAIWKGIAAGNFLEHTEFLGNIYGTSKTAVNTAAINNRICVMD
                     LNIDGVRSLKNTYLMPYSVYIRPTSLKMVETKLRRRNTEADDEIHRRVMLAKTDMDEA
                     NEAGLFDTIIIEDDVNLAYSKLIQILQDRIRMYFNTN"
     gene            6465. .7367
                     /gene="B1R"
     CDS             6465. .7367
                     /gene="B1R"
                     /codon_start=1
                     /protein_id="CAA75277.1"
                     /db_xref="GI:3097022"
                     /translation="MNFQGLVLTDNCKNQWVVGPLIGKGGFGSIYTTNDNNYVVKIEP
                     KANGSLFTEQAFYTRVLKPSVIEEWKKSHNIKHVGLITCKAFGLYKSINVEYRFLVIN
                     RLGADLDAVIRANNNRLPKRSVMLIGIEILNTIQFMHEQGYSHGDIKASNIVLDQIDK
                     NKLYLVDYGLVSKFMSNGEHVPFIRNPNKMDNGTLEFTPIDSHKGYVVSRRGDLETLG
                     YCMIRWLGGILPWTKISETKNCALVSATKQKYVNNTATLLMTSLQYAPRELLQYITMV
                     NSLTYFEEPNYDEFRHILMQGVYY"
```

```
gene            7458. .8969
                /gene="B2R"
CDS             7458. .8969
                /gene="B2R"
                /codon_start=1
                /protein_id="CAA75278.1"
                /db_xref="GI:3097023"
                /db_xref="SPTREMBL:O72739"
                /translation="MAMFYAHALGGYDENLHAFPGISSTVANDVRKYSVVSVYNNKYD
                IVKDKYMWCYSEVNKRYIGALLPMFECNEYLQIGDPIHDQEGNQISIITYRHKNYYAL
                SGIGYESLDLCLEGVGIHHHVLETGNAVYGKVQHDYSTIKEKAKEMSTLSPGPIIDYH
                VWIGDCICQVTAVDVHGKEIMKMRFKKGAVLQIPNLVKVKLGENDTENLSTAISALLN
                SGGGTIEVTSQEDRVNHVLMKRLESIRHMWSVVYDHFDIVNGKECCYVHMHSSNQNPI
                PSTVKTNLYMKTMGSCIQMDSMEALEYLSELKESGGRSPRPEMQEFEYPDGVEDTESI
                ERLAEEFFNRSELQAGESVKFGNSINVKHTSVSAKQLRTRIRQQLPSILSSFANTEGG
                YLFIGVDNNTHKVIGFTVGHDYLRLVENDIEKHIKRLRVVVHFCEKKEDIKYACRFIKV
                YKPGDETTSTYVCAIKVERCCCAVFADWPESWYMDTNGIKKYSPDEWVSHIKF"
gene            9194. .10870
                /gene="B3R"
CDS             9194. .10870
                /gene="B3R"
                /codon_start=1
                /protein_id="CAA75279.1"
                /db_xref="GI:3097024"
                /db_xref="SPTREMBL:O72740"
                /translation="MDFFKKEILDWSIYLSLHYIARACSNSSTSHIIQEYNLIRTYEK
                VDKTIVDFLSRWPNLFHILEYGENILHIYSMDDANANIIIFFLDRVLNINKNGSFIHN
                LGLSSSINIKEYVYQLVNNDHPDNRIRLMLENGRRTRHFLSYILDTVNIYISILINHG
                FYIDAEDSYGCTLLHRSIYHYKKSESESYNELIKILLNNGSDVDKKDTYGNTPFILLC
                KHDIDNAELFEICLENANIDSVDFNGYTPLHYVSCRNKYDFVKSLISKGANVNARNRF
                GTTPFYCGIIHGISLIKLYLESDTELEIDNEHIVRHLIIFDAVESLDYLLSRGVIDIN
                YRTIYSETSIYDAVSYNAYNTLVYLLNRNGDFETITTSGCTCISEAVANNNKIIMEVL
                LSKRPSLKIMIPSMIAITKHKQHNADLLKMCIKYTACMTDYDTLIDVQSLQQYKWYIL
                KCFDEIDIMKRCYIKNKTVFQLVFCIKDINTLMRYGKHPSFVKCARLDVYGSRVRNII
                ASIRYRQRLISLLSKKLDAGDKWSCFPNEIKYKILENFNDNELSTYLKIL"
gene            10974. .11927
                /gene="B4R"
CDS             10974. .11927
                /gene="B4R"
                /codon_start=1
                /protein_id="CAA75280.1"
                /db_xref="GI:3097025"
                /db_xref="SPTREMBL:O72741"
                /translation="MKTISVVTLLCVLPAVVYSTCTVPTMNNAKLTSTETSFNDKQKV
                TFTCDQGYHSLDPNAVCETDKWKYENPCKKMCTVSDYVSELYDKPLYEVNSTMTLSCN
                GETKYFRCEEKNGNTSWNDTVTCPNAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVG
                YEVIGASYISCTANSWNVIPSCQQKCDIPSLSNGLISGSTFSIGGVIHLSCKSGFTLT
                GSPSSTCIDGKWNPILPTCVRSNEEFDLVDDGPDDETDLSKISKDVVQYEQEIESLEA
                TYHIIIVALTIMGVIFLISIIVLVCSCDKNNDQYKFHKLLP"
gene            12023. .12574
                /gene="B5R"
CDS             12023. .12574
                /gene="B5R"
                /codon_start=1
                /protein_id="CAA75281.1"
                /db_xref="GI:3097026"
                /db_xref="SPTREMBL:O72742"
                /translation="MSSSVDVDIYDAVRAFLLRHYYDKRFIVYGRSNAILHNIYRLFT
                RCAVIPFDDIVRTMPNESRVKQWVIDTLNDIMTKEHDVSVSVGTGILFMEMFFDYNKN
                NSNSNSNSKNSINNQLMYDIINSVSIILANERYRSAFNDDGIYIRRTMIDKLYGYASL
                TTIGMITGGVCYYLLMHLVSLYK"
gene            12613. .13161
                /gene="B6R"
CDS             12613. .13161
                /gene="B6R"
                /codon_start=1
                /protein_id="CAA75282.1"
                /db_xref="GI:3097027"
                /db_xref="SPTREMBL:O72743"
                /translation="MYKNLITFLFVIGAVASYSNNEYTPFNKLSVKLYIDGVDNIENS
                YTDDNNELVLNFKEYTISIITESCDVGFDSIDIDVINDYKIIDMYTIDSSTIQRRGHT
                CKISTKLSCHYDKYPYIHKYDGDERQYSITAEGKCYKGIKYEISMMNDDTLLRKHTLK
                IGSTYIFDRHGHSNTYYSKYDF"
```

```
gene            13239. .14054
                /gene="B7R"

  Query Match          5.3%;  Score 42.8;  DB 16;  Length 49649;
  Best Local Similarity   54.4%;  Pred. No. 0.71;
  Matches   86;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     652 TTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCT 711
           || | || |||| ||    ||      | ||| |  | |      | |||||||  | |
Db   38470 TTCTCTGTAATTATGATTTAAACATTGATTATTGTATTCTTTTATGAAATTGTTGTTCAT 38411

Qy     712 TCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATT 771
           | ||||| ||  |||||   |||| ||||   | || |||   ||  | |  || || |
Db   38410 TAAGATGCAATAAGTTATTATATATCTTACGTAGTAATCAGTTGTAAGATATGGAATAAT 38351

Qy     772 ACTAATGAAACATTTAGAAAAACGCAATTATATCCATA 809
           |||||  ||  ||||    |     ||  ||||   ||
Db   38350 ACTAAGCAATGATTTGATTACGGATAAAGATATATCTA 38313


RESULT  11
AE001582/c
LOCUS       AE001582    18753 bp    DNA             BCT       16-NOV-1999
DEFINITION  Borrelia burgdorferi plasmid lp21, complete plasmid sequence.
ACCESSION   AE001582
VERSION     AE001582.1  GI:6382381
KEYWORDS    .
SOURCE      Lyme disease spirochete.
  ORGANISM  Borrelia burgdorferi
            Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1  (bases 1 to 18753)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
            White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
            van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
  TITLE     A bacterial genome in flux: The twelve linear and nine circular
            extrachromosomal DNAs in an infectious isolate of the Lyme disease
            spirochete Borrelia burgdorferi
  JOURNAL   Mol. Microbiol. (1999) In press
REFERENCE   2  (bases 1 to 18753)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
            White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
            van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .18753
                     /organism="Borrelia burgdorferi"
                     /db_xref="taxon:139"
                     /plasmid="lp21"
     gene            184. .618
                     /gene="BBU01"
     CDS             184. .618
                     /gene="BBU01"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07699.1"
                     /db_xref="GI:6382388"
                     /translation="MENYQYFFNLFFIFKIIFIYLYIMYKSVKEQQEKGIDHTCRILI
                     LTETIFEINLILENYSQKTLLKKYNENLKNKNLPPSNISTMKKYLNQLEKEIKIIAKF
                     YFKNDQSLIYYKLNYTLEKIWLKLIELFYKELKQFIQKNTTT"
     gene            746. .1114
                     /gene="BBU02"
     CDS             746. .1114
                     /gene="BBU02"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
```

/product="hypothetical protein"
/protein_id="AAF07700.1"
/db_xref="GI:6382389"
/translation="MDGVINDTLVARMKKQIKFNKNKLIILVKTLDHMNKELLYSANK
TYNYVLIQNNFNEALAKTYQLRVNYKTLLEYLEILEKNPKVILKRPTNKENESFIGLY
TLLYPLEDCCTKIYNSHPNI"
gene complement(1238. .1357)
/gene="BBU03"
CDS complement(1238. .1357)
/gene="BBU03"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF07701.1"
/db_xref="GI:6382390"
/translation="MVYYKLVRLYARIDVIISNEVGLREISLKEIVLSPIHLV"
gene 1486. .2628
/gene="BBU04"
CDS 1486. .2628
/gene="BBU04"
/note="similar to PID:1174341 PID:1174349 GB:AE001584
percent identity: 81.23; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07693.1"
/db_xref="GI:6382382"
/translation="MLISFVKESLPMHKIKTTNNNPRNCYNKVQYKLIVLISTICYLN
KTHKKYTQKTILYYFNENLRKNGQTISTLRTMQKYIYRLQKEIKVTTNYYQHMGVNSG
TEIYYKLNYPKKDCYHKINQHFKEKKETRFQNRVTNYFNKNSDSKMGSVQCESCNSNK
NNIKEERKINEIEKYQVINYFNKCNFSCKEILSILLNLNVDKDTMIKIIKTIKRTDIK
AKNKNIYFPKSCSKEKQEKLKKILCNTQKELEKSGYNSEQLETNFQKIYENYKYKPHF
IIENHKYSDLSYIKRKLEKSIERKKENSKQDYKNLRTNIFNILIEQLKKETNIEILKP
IIKEYLNNQKKIEYKKVFRIYYSELLEIINGKHYSNLKKFRKKSVG"
gene 2868. .3656
/gene="BBU05"
CDS 2868. .3656
/gene="BBU05"
/note="similar to GB:D26185 SP:P37522 PID:467381
PID:580906 GB:AL009126 percent identity: 55.98; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="plasmid partition protein, putative"
/protein_id="AAF07694.1"
/db_xref="GI:6382383"
/translation="MNKKPNILTIASLKGGVGKSTLTILFSYLLKDLGKKILIVDMDP
QNSITSYFSNYVHNVEIYNTYSMLKGGVSFNNCVGKINDHIFIIPAHPILENFNLEPI
DYKEIVLELRIEQNISAYEFDYILLDTPPHRDFLVKNALNITDHIIIPVQVERWSIES
FTILMNMIENFQIIKKKIYTTSIVENQFIKNRNTLKDIERLLYEKYGSYVKGKIHFSN
GIKVFINELIEPSSEESYYKEAKKTLHNILGYINHKNKYFIKNK"
gene 14633. .15235
/gene="BBU06"
CDS 14633. .15235
/gene="BBU06"
/note="similar to GB:U03641 PID:458218 percent identity:
61.64; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07695.1"
/db_xref="GI:6382384"
/translation="MEDIYMSKKNKKKLFLNDRIEHNIIDNLSNLDTNLDKNLMVYKD
LKEQLKLNLSNEIDTKIQRMRILYEIKTRKLYKYDGFKSFSQYLKTFVVAKTQAFFYL
KLYSKILEGSISIDTIKELGFENIKKHIFKKKSFDFNNNNKSKLDNKNISIRIFMRDK
ELYEFVKQDIKRASYIFKELFQNRKDVLTDIVTKYDNERN"
gene 15349. .15813
/gene="BBU07"
CDS 15349. .15813
/gene="BBU07"
/note="similar to PID:1174341 PID:1174349 GB:AE001584
percent identity: 81.71; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07696.1"
/db_xref="GI:6382385"
/translation="MIEILKTIKRTEIKAKNKNIHFTKSCSKEKQEKLKEILCNTQKE
LEKSGCNSEQLETNFQKIYENYKYKPHFIIENHKYSDLSYIKRKLEKSIEIKKENPQK
DYESLKINIFHIFIEQLKKEINIETLKPLVKEYLNNQKKIKYTKVFDTYYTR"
gene 15791. .16084
/gene="BBU08"
CDS 15791. .16084
/gene="BBU08"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF07702.1"
/db_xref="GI:6382391"
/translation="MIHITRGDNIKKYLNSHTKVLITKDTFKSYFFRFLWDKKNIVIF
DILKVYAQNLYKELINSEKTYNQVVVEIVENKELGYSINHRIIKVSSSKSIYY"
gene complement(16228. .16548)
/gene="BBU09"
CDS complement(16228. .16548)
/gene="BBU09"
/note="similar to GB:U03641 PID:458212 percent identity:
66.22; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07697.1"
/db_xref="GI:6382386"
/translation="MGIFYFQKECKTMSKAVDEIYCHSCGKTIKKDAEICISCGVRNK
QTENYNKLVVFLLCLFLGYLGVHRFYVGKIGTGMLYLFTFGFLYVGALIDLIRIATNK
FECK"
gene 16603. .16800
/gene="BBU10"
CDS 16603. .16800
/gene="BBU10"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF07703.1"
/db_xref="GI:6382392"
/translation="MVFLIEQLKKETNIEILKPIIKEYLNNQKKIKYNKIFGIYYLEL
LEIIKNRKHDLNLEKFSKKVV"
gene 16886. .17878
/gene="BBU11"
CDS 16886. .17878
/gene="BBU11"
/note="similar to SP:Q10618 PID:1322425 GB:AL123456
percent identity: 53.61; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07698.1"
/db_xref="GI:6382387"
/translation="MIKTPIISEDQFKIALKLIKMGEIIVFPTDTVYAIGANIYDEYA
VRMINLVKKKSINEPLMLYVDTIEKIKELSSHVPKSAKILMEKFSPGPLTYVLKKSIK
IPKFINNNLDTVAIRIPKNKTALNLIRKLKNPLVVSSANLSKRPSSTSFSMALKELNG
LVGGIIMPKKNKDSNIGIESTIIEFDSKDNVIILRPGAITKKMIEKVLGSRYKVNYAK
ETKIDLKKMFSYRLRIPVHTFKRGDNIRKYLKTNTKVLITRDTFKFYFFSFLWDKKNI
ILFDSLIEYAENLYKELVNAEEKYDQVIVEIVENEELGYAINNRIVKASSNNFM"
gene 17918. .18365
/gene="BBU12"
/note="This region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to

GB:L31615 GB:L31418 PID:520783 PID:551742 percent identity: 61.90; identified by sequence similarity; putative"

Query Match 5.3%; Score 42.6; DB 2; Length 18753;
Best Local Similarity 53.3%; Pred. No. 0.67;
Matches 90; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy    553 GTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGA 612
          |||| | |   | | |   | ||||||||| |  || ||||   |  |||||||  | ||
Db  16392 GTGGTATTTTTACTATGCTTATTTCTTGGTTATTTAGGAGTTCACAGATTTTATGTAGGT 16333

Qy    613 TACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTA 672
          |    |     ||| ||||| |  | |  ||  | || ||||     ||||     ||
Db  16332 AAAATAGGAACTGGTATGTTATACCTATTTACATTTGGATTTTTATATGTTGGAGCTTTA 16273

Qy    673 ATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAA 721
          || ||| | |||   ||  |||   | |   ||| |  |  | | | ||
Db  16272 ATCGATCTTATTAGAATAGCAACAAACAAGTTTGAATGTAAATAAGAAA 16224
```

RESULT 12
AB017069/c
LOCUS AB017069 88128 bp DNA PLN 20-NOV-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MNJ8, complete sequence.
ACCESSION AB017069
VERSION AB017069.1 GI:3510345
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MNJ8.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Nakamura,Y.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 88128)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935, Fax:+81-438-52-3934)
FEATURES Location/Qualifiers
source 1. .88128
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MNJ8"
/clone_lib="Mitsui P1"
BASE COUNT 29506 a 15101 c 14971 g 28550 t
ORIGIN

Query Match 5.1%; Score 41.6; DB 7; Length 88128;
Best Local Similarity 51.6%; Pred. No. 1.7;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy    348 ACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTAC 407
          ||||    |  |||  | | | | || |  |||     |||| | ||   | ||   |
Db  28303 ACAAATACATATTAATAAGCAAGATATGACGGACTAGTAACAAGTATCAAAGATCTAATC 28244

Qy    408 AAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAG 467
          |||| | ||   | |     |  |||||||| | | ||   | ||   ||||  |
Db  28243 AAACCATACTATTAACTAAGGCAATTTTCCAGTAACAAAACAATCTAAAAAGGCATGATT 28184

Qy    468 TGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCG 527
          |||    | ||||  | | |  |  ||   | ||| || ||| ||   |||||||   |
Db  28183 TGGACTAGTAACAAGTATCAAAGATAAAACAATTGCTTCTTTAAAAAACATATCTAAACC 28124

Qy    528 AAAT 531
          |||
Db  28123 CAAT 28120
```

RESULT 13
AF178650/c
LOCUS AF178650 73384 bp DNA PRI 02-SEP-1999
DEFINITION Homo sapiens chromosome 8 BAC clone 402H09, complete sequence.
ACCESSION AF178650
VERSION AF178650.1 GI:5817907
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 73384)
AUTHORS Blechschmidt,K., Schilhabel,M., Schattevoy,R., Baumgart,C., Menzel,U., Weber,J., Korenberg,J.R. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES Location/Qualifiers
source 1. .73384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="BAC 402H09"
gene <1. .33003
/gene="fertilin beta"
repeat_region 418. .469
/rpt_family="LTR16C"
repeat_region 606. .766
/rpt_family="HERVL"
repeat_region complement(2219. .2665)
/rpt_family="L1M4_orf2"
repeat_region complement(2632. .3027)
/rpt_family="L1M4_orf2"
repeat_region complement(3037. .3696)
/rpt_family="L1M4_orf2"
repeat_region complement(4110. .4176)
/rpt_family="MLT1J"
repeat_region 4602. .4677
/rpt_type=tandem
repeat_region complement(4665. .4800)
/rpt_family="L1PA5"
repeat_region 5072. .5097
/rpt_type=inverted
repeat_region complement(5863. .6031)
/rpt_family="L2"
repeat_region complement(6029. .6259)
/rpt_family="MIR"
repeat_region complement(6264. .6949)
/rpt_family="L2"
repeat_region 6463. .6488
/rpt_type=inverted
repeat_region 6499. .6810
/rpt_family="AluSg"
repeat_region 6781. .6816
/rpt_type=tandem
mRNA join(<7171. .7354,9495. .9593,9703. .9898,15453. .15558, 20824. .21007,26991. .27068,27289. .27427,30108. .30267, 31848. .31911,32677. .33003)
/gene="fertilin beta"
/note="mRNA acc.no. U52370"
/product="fertilin beta"
CDS join(<7171. .7354,9495. .9593,9703. .9898,15453. .15558, 20824. .21007,26991. .27068,27289. .27427,30108. .30267, 31848. .31881)
/gene="fertilin beta"

```
                /note="mRNA acc.no. U52370"
                /codon_start=2
                /product="fertilin beta"
                /protein_id="AAD52980.1"
                /db_xref="GI:5817908"
                /translation="HFSGVKIFSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVC
                GNAKLEAGEECDCGTEQDCALIGETCCDIATCRFKAGSNCAEGPCCENCLFMSKERMC
                RPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCMSGDKQCTDTFGKE
                VEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRAT
                IIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTD
                KCNDRGVCNNKKHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIE
                NIYHSKPMRWPFFLFIPFFIIFCVLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG
                "
exon            7171. .7354
                /gene="fertilin beta"
                /number=1
repeat_region   complement(7592. .7961)
                /rpt_family="L1MA9"
repeat_region   8816. .8849
                /rpt_type=tandem
exon            9495. .9593
                /gene="fertilin beta"
                /number=2
exon            9703. .9898
                /gene="fertilin beta"
                /number=3
repeat_region   10039. .10312
                /rpt_family="AluSg1"
repeat_region   10039. .10221
                /rpt_type=inverted
repeat_region   complement(11145. .11400)
                /rpt_family="L1M2_orf2"
repeat_region   complement(11478. .12025)
                /rpt_family="L1M4_5end"
repeat_region   11659. .11710
                /rpt_type=tandem
repeat_region   complement(11708. .12002)
                /rpt_family="AluSx"
repeat_region   11819. .12002
                /rpt_type=inverted
repeat_region   complement(12032. .12208)
                /rpt_family="LTR45"
repeat_region   complement(12344. .12610)
                /rpt_family="AluJo"
repeat_region   complement(12808. .12895)
                /rpt_family="L1M4_5end"
repeat_region   complement(13043. .13291)
                /rpt_family="LTR43"
repeat_region   13854. .13917
                /rpt_type=tandem
repeat_region   complement(13923. .14258)
                /rpt_family="MLT2E"
repeat_region   complement(14632. .14934)
                /rpt_family="AluY"
repeat_region   14752. .14940
                /rpt_type=inverted
exon            15453. .15558
                /gene="fertilin beta"
                /number=4
repeat_region   15796. .15851
                /rpt_type=tandem
repeat_region   complement(15823. .15899)
                /rpt_family="AluSg1"
repeat_region   15981. .16016
                /rpt_type=tandem
repeat_region   complement(15987. .16295)
                /rpt_family="AluY"
repeat_region   16298. .16535
                /rpt_family="MIR"
repeat_region   16563. .16749
                /rpt_type=inverted
repeat_region   16571. .16874
                /rpt_family="AluSx"
repeat_region   complement(16982. .17039)
                /rpt_family="L1MD3"
repeat_region   complement(17100. .17216)
                /rpt_family="L1MC4"
repeat_region   complement(17268. .17729)
                /rpt_family="L1MC4"
repeat_region   17348. .17387
                /rpt_type=tandem
repeat_region   complement(17720. .18144)
                /rpt_family="L1MC4"
repeat_region   18144. .18207
                /rpt_type=tandem
repeat_region   complement(18199. .18321)
                /rpt_family="L1MC4"
repeat_region   18315. .18350
                /rpt_type=tandem
repeat_region   complement(18442. .18492)
                /rpt_family="L1MC4"
repeat_region   complement(18750. .19124)
                /rpt_family="L1MEb_5end"
repeat_region   complement(19464. .20029)
                /rpt_family="MER90"
repeat_region   complement(20549. .20625)
                /rpt_family="MADE1"
repeat_region   20551. .20583
                /rpt_type=inverted
repeat_region   20593. .20625
                /rpt_type=inverted
exon            20824. .21007
                /gene="fertilin beta"
                /number=5
repeat_region   complement(21059. .21342)
                /rpt_family="AluSx"
repeat_region   22632. .22754
                /rpt_family="L1M4c_5end"
repeat_region   22656. .22753
                /rpt_family="L1M4_5end"
repeat_region   22888. .22921
                /rpt_type=tandem
repeat_region   complement(23000. .23371)
                /rpt_family="THE1-INTERNAL"
repeat_region   complement(23357. .23747)
                /rpt_family="THE1-INTERNAL"
repeat_region   complement(23751. .24076)
                /rpt_family="THE1B"
repeat_region   24689. .24895
                /rpt_family="AluSx"
repeat_region   25212. .25277
                /rpt_type=tandem
repeat_region   26680. .26839
                /rpt_family="L2a"
repeat_region   26728. .26839
                /rpt_family="L2b"
exon            26991. .27068
                /gene="fertilin beta"
                /number=6
repeat_region   27135. .27174
                /rpt_type=tandem
exon            27289. .27427
                /gene="fertilin beta"
                /number=7
repeat_region   complement(27944. .28069)
                /rpt_family="LTR33"
repeat_region   28650. .28742
```

```
  Query Match             5.1%;  Score 41.6;  DB 40;  Length 73384;
  Best Local Similarity   48.7%;  Pred. No. 1.6;
  Matches  113;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy    578 TTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGT 637
          ||  |||||     |  |||  |||| |  ||  |  ||||  ||  ||| ||  |||
Db   9077 TTTTATGGTATATTGTTGATTTATTTCATTTTTAACTCCCTAATTAAGGTCTGGAAAATG 9018
```

```
Qy     638 TTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGC 697
           |   | |   | |  | || | |||  | | | ||    |||| || ||| | | |
Db    9017 ATATTATTTCTGAGAATGCTGAAGTTGTCACCTTTATATCATTCACTCATATCTTATTAT 8958

Qy     698 AGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTA 757
           |||||  |   |    |  |  ||    || |  | ||   | |  |||     |||
Db    8957 GCCTTGTTTTATGCTTTAGTTCATCTATGCCCAGTGAATATGCTTTTAAAACTTCTATTA 8898

Qy     758 CAAGACTGAGTATTACTAATGAAACATTTAGAAAAACGCAATTATATCCATA 809
           | | | | ||||  |  |||  |  |    |  ||| |  | |  || || |
Db    8897 CTAAAGTTAGTAAAAGCAATAGATAAGAAGGCTAAAAGACAATGCATTCACA 8846
```

RESULT 14
BHV1CGEN
LOCUS BHV1CGEN 135301 bp DNA VRL 11-JAN-1999
DEFINITION Bovine herpesvirus 1 complete genome.
ACCESSION AJ004801
VERSION AJ004801.1 GI:2653291
KEYWORDS complete genome.
SOURCE Bovine herpesvirus type 1.1.
ORGANISM Bovine herpesvirus type 1.1
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Varicellovirus.
REFERENCE 1 (bases 99702 to 103033)
AUTHORS Wirth,U.V., Fraefel,C., Vogt,B., Vlcek,C., Paces,V. and Schwyzer,M.
TITLE Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' coterminal and encode a putative zinc finger transactivator protein
JOURNAL J. Virol. 66 (5), 2763-2772 (1992)
MEDLINE 92219360
REFERENCE 2 (bases 103034 to 111027; 127191 to 135300)
AUTHORS Schwyzer,M., Vlcek,C., Menekse,O., Fraefel,C. and Paces,V.
TITLE Promoter, spliced leader, and coding sequence for BICP4, the largest of the immediate-early proteins of bovine herpesvirus 1
JOURNAL Virology 197 (1), 349-357 (1993)
MEDLINE 94025583
REFERENCE 3 (bases 111028 to 114234; 123984 to 127190)
AUTHORS Schwyzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.
TITLE BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA which exhibits immediate early and late transcription kinetics
JOURNAL J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)
MEDLINE 94292919
REFERENCE 4 (bases 66900 to 96900)
AUTHORS Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D., Letchworth,G.J. and Schwyzer,M.
TITLE Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus 1 genome which exhibits a colinear gene arrangement with the UL21 to UL4 genes of herpes simplex virus
JOURNAL Virology 210 (1), 100-108 (1995)
MEDLINE 95313343
REFERENCE 5 (bases 1 to 31444)
AUTHORS Schwyzer,M., Styger,D., Vogt,B., Lowery,D.E., Simard,C., LaBoissiere,S., Misra,V., Vlcek,C. and Paces,V.
TITLE Gene contents in a 31-kb segment at the left genome end of bovine herpesvirus-1
JOURNAL Vet. Microbiol. 53 (1-2), 67-77 (1996)
MEDLINE 97164286
REMARK Accession# Z54206
REFERENCE 6 (bases 30801 to 67800)
AUTHORS Schwyzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G. and Misra,V.
TITLE Gene contents in a 37-kb segment centered in the UL part of the bovine herpesvirus 1 genome: the last gap
JOURNAL Unpublished
REMARK Accession# Z78205
REFERENCE 7 (bases 96901 to 99695)
AUTHORS Letchworth,G.J. and Kutish,G.F.
TITLE DNA sequence of the BHV-1 UL1 to UL3.5 genes
JOURNAL Unpublished
REFERENCE 8 (bases 114235 to 122983)
AUTHORS Goltz,M., Buhk,H.J., Broll,H., Lewin,M., Mankertz,A., Boerner,B., Borchers,K. and Weigelt,W.
TITLE Nucleotide sequence of the HindIII O and K fragments located in the US region of the bovine herpesvirus 1 genome
JOURNAL Unpublished
REMARK Accession# Z98199
REFERENCE 9 (bases 121402 to 123983)
AUTHORS Schwyzer,M.
TITLE Glycoprotein E and US9 genes of BHV1
JOURNAL Unpublished
REFERENCE 10 (bases 1 to 135301)
AUTHORS Schwyzer,M., Paces,V., Letchworth,G.J., Misra,V., Buhk,H.J., Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.
TITLE Complete DNA sequence of bovine herpesvirus 1
JOURNAL Unpublished
REFERENCE 11 (bases 1 to 135301)
AUTHORS Schwyzer,M.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1997) Institute of Virology, Faculty of Veterinary Medicine, University of Zurich, Winterthurerstrasse 266A, Zurich CH-8057, Switzerland
FEATURES Location/Qualifiers

source	1. .135301
	/organism="Bovine herpesvirus type 1.1"
	/db_xref="taxon:79889"
intron	<1. .472
	/note="3'part of intron spliced out from transcript over covalently joined genome ends, leading to immediate-early RNA 1.5 and circ protein"
source	order(1. .4014,102918. .111027,127191. .135300)
	/organism="Bovine herpesvirus type 1.2 (strain K22)"
	/virion
	/strain="K22"
	/db_xref="taxon:31519"
repeat_region	235. .430
	/rpt_unit=235. .248
	/rpt_type=DIRECT
mRNA	466. .1280
	/note="transcribed from alternative late promoter"
	/evidence=experimental
	/product="LR1.1 (circ)"
CDS	486. .1229
	/note="immediate-early and late myristylated protein (virion)"
	/codon_start=1
	/product="circ protein"
	/protein_id="CAA06078.1"
	/db_xref="GI:2653292"
	/db_xref="SPTREMBL:Q01342"
	/translation="MGARASAPAAGPPPAHAVLLDALSGGTIDLPGGDEAVFVSCPTT RPVYHHMRRGRTAHTTPVHFVGRAYAILPCRKFMLYLMRGGAVYGYEPTTGLHRLADS LHDFLTTAGLQQRDLHCLDVTVLDAQMDPVTFTTPEILIELEADPAFPPPPSARARRS TLRRASMRRPARTFCPHQLLAEGSILDLCSPEQAAAPGCSLLPACDSGDAACPCDAGE TARDCTADAARAPSPGALSRYSSVRSVFF"
polyA_signal	1259. .1264
	/gene="circ"
gene	1259. .1264
	/gene="circ"
polyA_signal	complement(1649. .1654)
	/gene="UL52"
gene	complement(1649. .7237)
	/gene="UL52"
gene	complement(1658. .2860)
	/gene="UL54"
CDS	complement(1658. .2860)
	/gene="UL54"
	/note="early BICP27 affecting mRNA processing (cell nucleus)"
	/codon_start=1
	/product="UL54 protein"
	/protein_id="CAA06079.1"
	/db_xref="GI:2653293"
	/db_xref="SPTREMBL:Q01354"
	/translation="MADPEIATLSTASESDDLSLFGSDREEDDEAPSLAPALRSVVGQ VRKRKLEGAEDEPMPAEPPGEGAASGDGGPAEAPPARRARVRPRRPRRRPRRRQPAGE

```
                QRSRGPAAKREAALATSSHGGGGAAARSIGSSLRLARSLAEAAQRATAERVTAVFAGA
                RLDLMRPVQNGGFRAAGVSPWAAVLDFGAEQFVPEGRRVTWETLMFHGRDLYRMFEVR
                PHAAQAARALRDLVLRSANLVDALASADECLTWCKFIATKNLRLRTKDPIVATAGAVL
                ENLRLKLAPFLRCYLRGRGLPSLEELCAARRLSLATCPASYMFVMLARLSRAVRSGAE
                CVPLLEVTVGDAPFEEYIPGTCVAGLIDALDTHKQACDSMTCKLVANFTLVPVYMHGK
                YFYCNEIF"
gene            complement(3040. .4038)
                /gene="UL53"
CDS             complement(3040. .4038)
                /gene="UL53"
                /codon_start=1
                /product="glycoprotein K"
                /protein_id="CAA06080.1"
                /db_xref="GI:2653294"
                /db_xref="SPTREMBL:Q65816"
                /translation="MLLGGRTVNLAALALLTAHLALALWVALAARCQRCACVRATARN
                GSLRWELRSPGAVYVWGGANNATLAADAPCRHAVVQHIPPGLLDGDEALHGRVRAVAG
                ARDCRAYLWCAQARGGLLAWLLYVAFVYLRQERRMFGLCRNDADFLSPGGYTLNYAAA
                ALAAVVGHGPYTKLARLMCELSARRRALAVDFRLDPLGCAWRPRAALPLLAEGFARLG
                ARIAAAGSVGITHPCAAAYPLYLKIWAWVHVALFAGLELVSLLYRKPRRRGGTCAGDG
                GDGGESGIRKVCVNCCSTLLAGLLVKALYLAAIVGGVIALLHYEHNLRLRLLGAQT"
CDS             complement(4013. .7237)
                /gene="UL52"
                /codon_start=1
                /product="component of DNA helicase/primase complex"
                /protein_id="CAA06081.1"
                /db_xref="GI:2653295"
                /db_xref="SPTREMBL:Q65817"
                /translation="MDGPAALEDALQILYATDGCAIGFSLMLLTGQEEPGRGAYVVSY
                NWPGRLAACLGGESPNDVALALADARPLVAFYLLGGPPGGSAAATPRLARLRVHLQQR
                AGRARARRRTAPRQGPTRARAARAPRRGRHLCAARGHDHALLVATEQLAPRTGRTADD
                ARYVEGQTTVRSALRARPAGRRGLASLYIHHEHKTVAAYRRLYSNSGATPFWFLSKFG
                PGEKTLVLATRFYVFQAERAGDAATYDLQAVRDCLATYAVAAPPNPSGLAFPDLVSFA
                ALAAFCCRSGYARGAVAAGAPAYVAARIEADLAEVRCLREYIDHDRRSLKVADPEFVA
                YVYLAYFEGFNRRQITEHLRAVTATEPPDGGEALAPRLAGVSRLRERAVDAFFRHVRA
                QFNVQSYIEQNVAVAVVRLPPATAEAYARARTYARLLAATANGPGRTICDGAAALRAA
                LDHLEGQAARFGWVLHAPGAPSASAVPGPGLDAALRVGAPDGGASPDATPACCGVSKR
                LLELAAAAAARDGAHPLDGLFGARGTAAPTPVYRVELPRGHQAFAVADGDDWAAVTSVA
                ALDPDEVAVEVAAAAAAAAESGRDALVERDARLTALLTERAACAAGRSDRRSLGPAAP
                RDQYYVNRNELFNARLAVTNIVLDVDFRLKRPLPRGDLHGAMRSFRRGALAALAALFP
                EIAADAWPAHPCFFYKSACPSAAASNLGDGDASPSWAAPDRDAGWDEYESALAREGEA
                QYSCEDVSDPAGARPEETGGAEGAGSAAALTTACGCDDKMGFRVAVPVPAPYIIAGGP
                TLKGIARLVQHAVMLERPFAEAMSRYLRDFAFVDAAVYTHGHSLRLPFFGKPDAASNV
                GRWLLPFYVVPERCDDVAAFVAAHRDPHHFHFHAAPPAGGRREARRAHVGGEYVSFFE
                RKAAVNRAAFHGVRVSLRRALAAVGVDGGDRGAVEEFIADVIEAELAPYMAEHYPQAA
                REYQGAGARLVAAKADWLLAQLVPARAGTAQGFGCLRAAHGRAAQNKARSLVSLSVDS
                HDRLCASLIQQCFATKCGSNRLGTVFTVDLSGRAARYAARGADC"
source          order(4015. .9664,15010. .22094,26636. .102917,
                114235. .123983,135301)
                /organism="Bovine herpesvirus type 1.1 (strain Cooper)"
                /virion
                /strain="Cooper"
                /db_xref="taxon:10323"
gene            7236. .7981
                /gene="UL51"
CDS             7236. .7967
                /gene="UL51"
                /codon_start=1
                /product="virion protein"
                /protein_id="CAA06082.1"
                /db_xref="GI:2653296"
                /db_xref="SPTREMBL:Q65818"
                /translation="MALQRLVGWLCGRGGGTRDAPQYEPLARRLSGPTLFRLQEAVVA
                VSALLPAPLTVEDVARSADGTRRLAKAQSLARTYYICQRNIECLSKHQAACSDASITA
                VVTKHIQDAQRMRDTCLAALLQMYHSVGAVEGTTDSMVDQAIRMAAESNIVMADVAVL
                ERALGIQAQGAGAEAAARASPTAVGAEARAPPPLPRQPAASAPVAVSPAPPVAPAEAA
                ATKPPSRAGRAAAAASSPAVLQLAA"
polyA_signal    7976. .7981
                /gene="UL51"

Query Match             4.9%;  Score 39.8;  DB 16;  Length 135301;
Best Local Similarity   48.1%;  Pred. No. 5.5;
Matches  113;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy      76 GGCCTCTATCCTATGCGAGGTCCCTTTAAGAACCTCGCCCTGTTGCCCTTCTCCCTCCCG 135
           ||| |  ||  ||| || ||     | | |  || | |  ||| |  | | ||||
Db 106797 GGCGGCGGCCCGTCGCGCGGCGCCGCGGCGTAGCCAGCGCGGGCGCCGGTTGCGCGCCCG 106856

Qy     136 CTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAGATGGCGGCC 195
           | |  | || | |||||  | ||    | |      |    || | ||  |    ||| |
Db 106857 CGCTGGCGCCGCGGCGGGGGGGGCGCCGTCTCCGGCGGCGAGGGCGCCGGGGGCCCGGGC 106916

Qy     196 GCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTG 255
           ||  ||||    ||  || | |  |||| | |||   |   | |   || |   || |||
Db 106917 GCGCGGCCCCGCGGGGCGCCGGGCCCGGCGCCCGCCTCTCGCGGGGCCGCTTTGGTTCTG 106976

Qy     256 TGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGG 310
           | | ||||| | |   | ||| ||   | |   |  | || ||   ||
Db 106977 GTCCCGGCCCCAGTCCCGGCCTCAGCCTCGGCCTCGGCCTCAGCGTCGGCGTCGG 107031


RESULT 15
BHV1CGEN/c
LOCUS       BHV1CGEN   135301 bp    DNA             VRL       11-JAN-1999
DEFINITION  Bovine herpesvirus 1 complete genome.
ACCESSION   AJ004801
VERSION     AJ004801.1  GI:2653291
KEYWORDS    complete genome.
SOURCE      Bovine herpesvirus type 1.1.
  ORGANISM  Bovine herpesvirus type 1.1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 99702 to 103033)
  AUTHORS   Wirth,U.V., Fraefel,C., Vogt,B., Vlcek,C., Paces,V. and Schwyzer,M.
  TITLE     Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
            are 3' coterminal and encode a putative zinc finger transactivator
            protein
  JOURNAL   J. Virol. 66 (5), 2763-2772 (1992)
  MEDLINE   92219360
REFERENCE   2  (bases 103034 to 111027; 127191 to 135300)
  AUTHORS   Schwyzer,M., Vlcek,C., Menekse,O., Fraefel,C. and Paces,V.
  TITLE     Promoter, spliced leader, and coding sequence for BICP4, the
            largest of the immediate-early proteins of bovine herpesvirus 1
  JOURNAL   Virology 197 (1), 349-357 (1993)
  MEDLINE   94025583
REFERENCE   3  (bases 111028 to 114234; 123984 to 127190)
  AUTHORS   Schwyzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.
  TITLE     BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA
            which exhibits immediate early and late transcription kinetics
  JOURNAL   J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)
  MEDLINE   94292919
REFERENCE   4  (bases 66900 to 96900)
  AUTHORS   Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D.,
            Letchworth,G.J. and Schwyzer,M.
  TITLE     Nucleotide sequence analysis of a 30-kb region of the bovine
            herpesvirus 1 genome which exhibits a colinear gene arrangement
            with the UL21 to UL4 genes of herpes simplex virus
  JOURNAL   Virology 210 (1), 100-108 (1995)
  MEDLINE   95313343
REFERENCE   5  (bases 1 to 31444)
  AUTHORS   Schwyzer,M., Stygen D., Vogt,B., Lowery,D.E., Simard,C.,
            LaBoissiere,S., Misra,V., Vlcek,C. and Paces,V.
  TITLE     Gene contents in a 31-kb segment at the left genome end of bovine
            herpesvirus-1
  JOURNAL   Vet. Microbiol. 53 (1-2), 67-77 (1996)
  MEDLINE   97164286
  REMARK    Accession# Z54206
REFERENCE   6  (bases 30801 to 67800)
  AUTHORS   Schwyzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G. and
            Misra,V.
  TITLE     Gene contents in a 37-kb segment centered in the UL part of the
            bovine herpesvirus 1 genome: the last gap
  JOURNAL   Unpublished
  REMARK    Accession# Z78205
REFERENCE   7  (bases 96901 to 99695)
  AUTHORS   Letchworth,G.J. and Kutish,G.F.
```

```
  TITLE     DNA sequence of the BHV-1 UL1 to UL3.5 genes
  JOURNAL   Unpublished
REFERENCE   8  (bases 114235 to 122983)
  AUTHORS   Goltz,M., Buhk,H.J., Broll,H., Lewin,M., Mankertz,A., Boerner,B.,
            Borchers,K. and Weigelt,W.
  TITLE     Nucleotide sequence of the HindIII O and K fragments located in the
            US region of the bovine herpesvirus 1 genome
  JOURNAL   Unpublished
  REMARK    Accession# Z98199
REFERENCE   9  (bases 121402 to 123983)
  AUTHORS   Schwyzer,M.
  TITLE     Glycoprotein E and US9 genes of BHV1
  JOURNAL   Unpublished
REFERENCE   10 (bases 1 to 135301)
  AUTHORS   Schwyzer,M., Paces,V., Letchworth,G.J., Misra,V., Buhk,H.J.,
            Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.
  TITLE     Complete DNA sequence of bovine herpesvirus 1
  JOURNAL   Unpublished
REFERENCE   11 (bases 1 to 135301)
  AUTHORS   Schwyzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1997) Institute of Virology, Faculty of
            Veterinary Medicine, University of Zurich, Winterthurerstrasse
            266A, Zurich CH-8057, Switzerland
FEATURES             Location/Qualifiers
     source          1. .135301
                     /organism="Bovine herpesvirus type 1.1"
                     /db_xref="taxon:79889"
     intron          <1. .472
                     /note="3'part of intron spliced out from transcript over
                     covalently joined genome ends, leading to immediate-early
                     RNA 1.5 and circ protein"
     source          order(1. .4014,102918. .111027,127191. .135300)
                     /organism="Bovine herpesvirus type 1.2 (strain K22)"
                     /virion
                     /strain="K22"
                     /db_xref="taxon:31519"
     repeat_region   235. .430
                     /rpt_unit=235. .248
                     /rpt_type=DIRECT
     mRNA            466. .1280
                     /note="transcribed from alternative late promoter"
                     /evidence=experimental
                     /product="LR1.1 (circ)"
     CDS             486. .1229
                     /note="immediate-early and late myristylated protein
                     (virion)"
                     /codon_start=1
                     /product="circ protein"
                     /protein_id="CAA06078.1"
                     /db_xref="GI:2653292"
                     /db_xref="SPTREMBL:Q01342"
                     /translation="MGARASAPAAGPPPAHAVLLDALSGGTIDLPGGDEAVFVSCPTT
                     RPVYHHMRRGRTAHTTPVHFVGRAYAILPCRKFMLYLMRGGAVYGYEPTTGLHRLADS
                     LHDFLTTAGLQQRDLHCLDVTVLDAQMDPVTFTTPEILIELEADPAFPPPPSARARRS
                     TLRRASMRRPARTFCPHQLLAEGSILDLCSPEQAAAPGCSLLPACDSGDAACPCDAGE
                     TARDCTADAARAPSPGALSRYSSVRSVFF"
     polyA_signal    1259. .1264
                     /gene="circ"
     gene            1259. .1264
                     /gene="circ"
     polyA_signal    complement(1649. .1654)
                     /gene="UL52"
     gene            complement(1649. .7237)
                     /gene="UL52"
     gene            complement(1658. .2860)
                     /gene="UL54"
     CDS             complement(1658. .2860)
                     /gene="UL54"
                     /note="early BICP27 affecting mRNA processing (cell
                     nucleus)"
                     /codon_start=1
                     /product="UL54 protein"
                     /protein_id="CAA06079.1"
                     /db_xref="GI:2653293"
                     /db_xref="SPTREMBL:Q01354"
                     /translation="MADPEIATLSTASESDDLSLFGSDREEDDEAPSLAPALRSVVGQ
                     VRKRKLEGAEDEPMPAEPPGEGAASGDGGPAEAPPARRARVRPRRPRRRPRRRQPAGE
                     QRSRGPAAKREAALATSSHGGGGAAARSIGSSLRLARSLAEAAQRATAERVTAVFAGA
                     RLDLMRPVQNGGFRAAGVSPWAAVLDFGAEQFVPEGRRVTWETLMFHGRDLYRMFEVR
                     PHAAQAARALRDLVLRSANLVDALASADECLTWCKFIATKNLRLRTKDPIVATAGAVL
                     ENLRLKLAPFLRCYLRGRGLPSLEELCAARRLSLATCPASYMFVMLARLSRAVRSGAE
                     CVPLLEVTVGDAPFEEYIPGTCVAGLIDALDTHKQACDSMTCKLVANFTLVPVYMHGK
                     YFYCNEIF"
     gene            complement(3040. .4038)
                     /gene="UL53"
     CDS             complement(3040. .4038)
                     /gene="UL53"
                     /codon_start=1
                     /product="glycoprotein K"
                     /protein_id="CAA06080.1"
                     /db_xref="GI:2653294"
                     /db_xref="SPTREMBL:Q65816"
                     /translation="MLLGGRTVNLAALALLTAHLALALWVALAARCQRCACVRATARN
                     GSLRWELRSPGAVYVWGGANNATLAADAPCRHAVVQHIPPGLLDGDEALHGRVRAVAG
                     ARDCRAYLWCAQARGGLLAWLLYVAFVYLRQERRMFGLCRNDADFLSPGGYTLNYAAA
                     ALAAVVGHGPYTKLARLMCELSARRRALAVDFRLDPLGCAWRPRAALPLLAEGFARLG
                     ARIAAAGSVGITHPCAAAYPLYLKIWAWVHVALFAGLELVSLLYRKPRRRGGTCAGDG
                     GDGGESGIRKVCVNCCSTLLAGLLVKALYLAAIVGGVIALLHYEHNLRLRLLGAQT"
     CDS             complement(4013. .7237)
                     /gene="UL52"
                     /codon_start=1
                     /product="component of DNA helicase/primase complex"
                     /protein_id="CAA06081.1"
                     /db_xref="GI:2653295"
                     /db_xref="SPTREMBL:Q65817"
                     /translation="MDGPAALEDALQILYATDGCAIGFSLMLLTGQEEPGRGAYVVSY
                     NWPGRLAACLGGESPNDVALALADARPLVAFYLLGGPPGGSAAATPRLARLRVHLQQR
                     AGRARARRRTAPRQGPTRARAARAPRRGRHLCAARGHDHALLVATEQLAPRTGRTADD
                     ARYVEGQTTVRSALRARPAGRRGLASLYIHHEHKTVAAYRRLYSNSGATPFWFLSKFG
                     PGEKTLVLATRFYVFQAERAGDAATYDLQAVRDCLATYAVAAPPNPSGLAFPDLVSFA
                     ALAAFCCRSGYARGAVAAGAPAYVAARIEADLAEVRCLREYIDHDRRSLKVADPEFVA
                     YVYLAYFEGFNRRQITEHLRAVTATEPPDGGEALAPRLAGVSRLRERAVDAFFRHVRA
                     QFNVQSYIEQNVAVAVVRLPPATAEAYARARTYARLLAATANGPGRTICDGAAALRAA
                     LDHLEGQAARFGWVLHAPGAPSASAVPGPGLDAALRVGAPDGGASPDATPACCGVSKR
                     LLELAAAAARDGAHPLDGLFGARGTAAPTPVYRVELPRGHQAFAVADGDDWAAVTSVA
                     ALDPDEVAVEVAAAAAAAAESGRDALVERDARLTALLTERAACAAGRSDRRSLGPAAP
                     RDQYYVNRNELFNARLAVTNIVLDVDFRLKRPLPRGDLHGAMRSFRRGALAALAALFP
                     EIAADAWPAHPCFFYKSACPSAAASNLGDGDASPSWAAPDRDAGWDEYESALAREGEA
                     QYSCEDVSDPAGARPEETGGAEGAGSAAALTTACGCDDKMGFRVAVPVPAPYIIAGGP
                     TLKGIARLVQHAVMLERPFAEAMSRYLRDFAFVDAAVYTHGHSLRLPFFGKPDAASNV
                     GRWLLPFYVVPERCDDVAAFVAAHRDPHHFHFHAAPPAGGRREARRAHVGGEYVSFFE
                     RKAAVNRAAFHGVRVSLRRALAAVGVDGGDRGAVEEFIADVIEAELAPYMAEHYPQAA
                     REYQGAGARLVAAKADWLLAQLVPARAGTAQGFGCLRAAHGRAAQNKARSLVSLSVDS
                     HDRLCASLIQQCFATKCGSNRLGTVFTVDLSGRAARYAARGADC"
     source          order(4015. .9664,15010. .22094,26636. .102917,
                     114235. .123983,135301)
                     /organism="Bovine herpesvirus type 1.1 (strain Cooper)"
                     /virion
                     /strain="Cooper"
                     /db_xref="taxon:10323"
     gene            7236. .7981
                     /gene="UL50"
     CDS             7236. .7967
                     /gene="UL51"
                     /codon_start=1
                     /product="virion protein"
                     /protein_id="CAA06082.1"
                     /db_xref="GI:2653296"
                     /db_xref="SPTREMBL:Q65818"
                     /translation="MALQRLVGWLCGRGGGTRDAPQYEPLARRLSGPTLFRLQEAVVA
                     VSALLPAPLTVEDVARSADGTRRLAKAQSLARTYYICQRNIECLSKHQAACSDASITA
                     VVTKHIQDAQRMRDTCLAALLQMYHSVGAVEGTTDSMVDQAIRMAAESNIVMADVAVL
                     ERALGIQAQGAGAEAAARASPTAVGAEARAPPPLPRQPAASAPVAVSPAPPVAPAEAA
                     ATKPPSRAGRAAAAASSPAVLQLAA"
     polyA_signal    7976. .7981
```

/gene="UL51"

Query Match 4.9%; Score 39.8; DB 16; Length 135301;
Best Local Similarity 48.1%; Pred. No. 5.5;
Matches 113; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

```
Qy      76 GGCCTCTATCCTATGCGAGGTCCCTTTAAGAACCTCGCCCTGTTGCCCTTCTCCCTCCCG 135
           |||  |   ||   ||| ||  ||     | | |  || | |  |||  |  | | ||||
Db 131421 GGCGGCGGCCCGTCGCGCGGCGCCGCGGCGTAGCCAGCGCGGGCGCCGGTTGCGCGCCCG 131362

Qy     136 CTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAGATGGCGGCC 195
           | |  | || | |||||  | ||    | |      |    || | ||  |    ||| |
Db 131361 CGCTGGCGCCGCGGCGGGGGGGGCGCCGTCTCCGGCGGCGAGGGCGCCGGGGGCCCGGGC 131302

Qy     196 GCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTG 255
           ||  ||||    ||  || | |  |||| | |||   |   | |   || |   || |||
Db 131301 GCGCGGCCCCGCGGGGCGCCGGGCCCGGCGCCCGCCTCTCGCGGGGCCGCTTTGGTTCTG 131242

Qy     256 TGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGG 310
             | | |||| |  |   | ||| ||   | |   |  | || ||   ||
Db 131241 GTCCCGGCCCCAGTCCCGGCCTCAGCCTCGGCCTCGGCCTCAGCGTCGGCGTCGG 131187
```

Search completed: February 11, 2000, 01:45:19
Job time: 3932 sec

OM protein - protein search, using sw model

Run on: February 11, 2000, 05:02:22 ; Search time 35.19 Seconds
(without alignments)
181.062 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 269
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: OLIGO

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

[redacted]d size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	269	100.0	269	1	W94291	Human beta-amyloid
2	133	49.4	139	1	Y12358	Human 5' EST secre
3	82	30.5	148	1	Y12426	Human 5' EST secre
4	8	3.0	757	1	W71297	Lucilia cuprina ec
5	7	2.6	283	1	R03623	Zucchini yellow mo
6	7	2.6	828	1	R07661	Incomplete form of
7	7	2.6	979	1	R08338	Complete form of h
8	7	2.6	437	1	R08390	Truncated form of
9	7	2.6	236	1	R14702	Vh-Lab-Vl construc
10	7	2.6	223	1	R20227	pY53.3 (NCIMB 4030
11	7	2.6	461	1	R24048	Heavy chain of 4D5
12	7	2.6	484	1	R27097	Translation of the
13	7	2.6	3080	1	R35081	ZYMV polyprotein.
14	7	2.6	211	1	R50174	Membrane anchor pe
15	7	2.6	211	1	R54281	M13 phage coat pro
16	7	2.6	302	1	R60206	Bispecific CD3-L6F
17	7	2.6	452	1	R59792	Aspergillus tubige
18	7	2.6	211	1	R62927	Bacteriophage coat
19	7	2.6	293	1	R75660	hEGF/M13 pIII fusi
20	7	2.6	246	1	R79869	Anti-EGFR single c
21	7	2.6	2713	1	W00168	Human T cell inosi
22	7	2.6	2695	1	W07632	Human type I inosi
23	7	2.6	399	1	W06322	Human mitogen-acti
24	7	2.6	233	1	W19806	phoA signal:KPI(1-
25	7	2.6	232	1	W19807	Protein encoded by
26	7	2.6	233	1	W19808	phoA:KIP(1-55;M15A
27	7	2.6	46	1	W24420	Nucleic acid deliv
28	7	2.6	794	1	W22720	Human serum albumi
29	7	2.6	784	1	W22717	Human serum albumi
30	7	2.6	789	1	W22718	Human serum albumi
31	7	2.6	534	1	W25769	Human MLN 51. Gene
32	7	2.6	240	1	W25784	Anti-gp130 antibod
33	7	2.6	924	1	W69848	Amino acid sequenc
34	7	2.6	745	1	W33810	Microscilla furves
35	7	2.6	412	1	W70672	Protein encoded by
36	7	2.6	432	1	W82246	Bacteriophage fd-t
37	7	2.6	698	1	W83493	4D5 Fab molecule e
38	7	2.6	399	1	W97672	Human mitogen acti
39	7	2.6	396	1	W93966	Plasmid pBBP20 pro
40	6	2.2	293	1	P91368	45 kDa amino termi
41	6	2.2	80	1	P92067	Partial MF alpha-1
42	6	2.2	20	1	R03360	Antigenic fragment
43	6	2.2	24	1	R03361	Antigenic fragment
44	6	2.2	25	1	R03362	Antigenic fragment
45	6	2.2	82	1	W99773	DEF chimeric molec

ALIGNMENTS

RESULT 1
W94291
ID W94291 standard; Protein; 269 AA.
AC W94291;
DT 27-APR-1999 (first entry)
DE Human beta-amyloid peptide-binding protein (BBP).
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;
KW human; Alzheimer's disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 68. .269
FT /note= "specifically claimed fragment having
FT beta-amyloid peptide binding activity"
PN WO9846636-A2.
PD 22-OCT-1998.
PF 14-APR-1998; U07462.
PR 16-APR-1997; US-064583.
PA (AMHP) AMERICAN HOME PROD CORP.
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,
PI Walker SG;
DR WPI; 99-080736/07.
DR N-PSDB; X05735.
PT Polynucleotide encoding beta-amyloid peptide binding protein - used
PT to identify inhibitors of beta-amyloid peptide for treating
PT Alzheimer's disease
PS Claim 7; Pages 43-44; 59pp; English.
CC The present sequence represents a beta-amyloid peptide binding protein
CC (BBP). The polynucleotide comprising the entire BBP nucleotide sequence
CC of clone BBP1-f1 is deposited under the accession number ATCC 98617. The
CC polynucleotide comprising a fragment of BBP (nucleotides 202-807 of the
CC full length BBP) of clone pEK196 is deposited as ATCC 98399. Host cells
CC transformed with a vector comprising the BBP nucleic acid are used for
CC the recombinant production of the protein. The protein can be used in a
CC method for diagnosing a disease characterised by aberrant expression of
CC human beta-amyloid protein (BAP). The protein can also be used in a
CC method for screening for compounds which regulate expression of a BAP
CC binding protein. The proteins, antibodies and identified compounds can be
CC used in the treatment or prevention of Alzheimer's disease.
SQ Sequence 269 AA;

Query Match 100.0%; Score 269; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.4e-239;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60

Qy      61 SKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC 120

Qy     121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180

Qy     181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240

Qy     241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269
```

```
          ||||||||||||||||||||||||||||
Db    241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269


RESULT   2
Y12358
ID   Y12358 standard; Protein; 139 AA.
AC   Y12358;
DT   17-JUN-1999  (first entry)
DE   Human 5' EST secreted protein SEQ ID NO:389.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   N-PSDB; X41191.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 27; Page 714-715; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   139 AA;


  Query Match             49.4%;  Score 133;  DB 1;  Length 139;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-115;
  Matches  133;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 SGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDA 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7 SGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDA 66

Qy     129 TQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 TQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 126

Qy     189 LSLFLGWLGADRF 201
           |||||||||||||
Db     127 LSLFLGWLGADRF 139


RESULT   3
Y12426
ID   Y12426 standard; Protein; 148 AA.
AC   Y12426;
DT   17-JUN-1999  (first entry)
DE   Human 5' EST secreted protein SEQ ID NO:457.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   N-PSDB; X41259.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 27; Page 763-764; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   148 AA;


  Query Match             30.5%;  Score 82;  DB 1;  Length 148;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-68;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      73 APEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEP 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 APEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEP 70

Qy     133 VNCTNYTAHVSCFPAPNITCKD 154
           ||||||||||||||||||||||
Db      71 VNCTNYTAHVSCFPAPNITCKD 92


RESULT   4
W71297
ID   W71297 standard; Protein; 757 AA.
AC   W71297;
DT   02-FEB-1999  (first entry)
DE   Lucilia cuprina ecdysteroid receptor.
KW   Ecdysteroid receptor; insecticide; biological control;
KW   gene therapy; Australian blowfly.
OS   Lucilia cuprina.
FH   Key             Location/Qualifiers
FT   Domain          1. .300
FT                   /label= A/B
FT   Domain          301. .366
FT                   /label= C
FT                   /note= "DNA binding domain"
FT   Domain          367. .453
FT                   /label= D
FT   Domain          545. .674
FT                   /label= E
FT                   /note= "hormone binding domain"
FT   Domain          675. .757
FT                   /label= F
```

FT Peptide 379. .385
FT /label= NLS
FT /note= "nuclear localisation signal"
FT Peptide 387. .391
FT /label= NLS
FT /note= "nuclear localisation signal"
FT Region 495. .536
FT /note= "helix-turn-zipper motif"
PN WO9835550-A2.
PD 20-AUG-1998.
PF 16-FEB-1998; NZ0018.
PR 14-FEB-1997; NZ-314239.
PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
PA (NZWO-) NEW ZEALAND WOOL BOARD.
PI Heath ACG, Huang SQ;
DR WPI; 98-520754/44.
DR N-PSDB; V60302.
PT New Lucilia cuprina ecdysteroid receptor, its fragments and related
PT nucleic acid - vectors, transformed cells, ligands and antisense
PT sequences, used as insecticides also in inducible gene therapy
PT systems and for recombinant protein production
Example 2; Fig 4; 50pp; English.
CC This is the amino acid sequence of novel claimed Lucilia cuprina
CC ecdysteroid receptor (EcR), as deduced from the sequence of an
CC isolated polynucleotide (see V60302). Also claimed are peptides
CC containing one or more of the A/B, C, D, E or F (especially C or E)
CC domains of the ecdysteroid receptor, nucleic acid encoding them,
CC expression vectors, an inducible expression system, ligands able to
CC bind the receptor or its peptides, and a vector encoding antisense
CC RNA. Vectors that express the receptor, or its C or E domains,
CC vectors containing antisense sequences, and vectors in which a gene
CC encoding the ecdysteroid receptor is inactivated are used as
CC insecticides, particularly against flies and specifically L. cuprina
CC (Australian blowfly, a vector of human and animal diseases).
CC Overexpressing the receptor induces premature moulting while
CC antisense RNA prevents moulting. Ligands can also be used as
CC insecticides and vectors are particularly used in baits for adult
CC flies. Also contemplated is germline transformation of flies using
CC a transposon encoding the receptor or its C or E domains. Host
CC cells are used to produce recombinant receptor, and inducible
CC expression systems are used for steroid-responsive expression of
CC other genes, e.g. in gene therapy (where expression will be induced
CC by administration of insect steroids) or for large scale protein
CC production.
SQ Sequence 757 AA;

Query Match 3.0%; Score 8; DB 1; Length 757;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 LGGGGSGS 54
||||||||
Db 147 LGGGGSGS 154

RESULT 5
R03623
ID R03623 standard; protein; 283 AA.
AC R03623;
DT 07-AUG-1990 (first entry)
DE Zucchini yellow mosaic virus (ZYMV) coat protein.
KW ZYMV; curcurbitaceae; solanaceae; leguminosea; caricacea;
KW potyvirus; ds.
OS Papaya ringspot virus.
PN WO9002184-A.
PD 8-MAR-1990.
PF 20-JUL-1989; 003094.
PR 19-AUG-1988; US-234412.
PR 14-MAR-1989; US-323536.
PR 19-JUN-1989; US-368710.
PA (CORR) Cornell Res Found Inc, (UPJO) The Upjohn Company.
PI Quemada H, Slightom JL, Gonsalves D, L'Hostis B;
DR WPI; 90-099408/13.
DR N-PSDB; Q03671.
PT Potyvirus coat protein genes - used to produce transformed plants
PT resistant to viral infection by potyvirus and related viruses.
PS Disclosure; p; English.
CC Plant cells may transformed with coat protein to render them resistant
CC to attack by potyvirus and related viruses from whence the sequence is
CC derived - papaya ringspot virus, Watermelon mosaic virus II and Zucchini
CC yellow mosaic virus.
CC The promoter and polyadenylation signals are from the Cauliflower mosaic
CC virus 35s promoter, and initiation region is derived from the Cucumber
CC mosaic virus coat protein or RUBISCO gene.
SQ Sequence 283 AA;

Query Match 2.6%; Score 7; DB 1; Length 283;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 GSGSGEK 57
|||||||
Db 32 GSGSGEK 38

RESULT 6
R07661
ID R07661 standard; protein; 828 AA.
AC R07661;
DT 25-FEB-1991 (first entry)
DE Incomplete form of human RF-X protein.
KW MHC class II; down regulation; autoimmune disease.
OS Homo sapiens.
PN WO9012812-A.
PD 01-NOV-1990.
PF 18-APR-1990; E00625.
PR 18-APR-1989; EP-106944.
PR 14-AUG-1989; EP-115008.
PA (MACH/) Mach B.
PI Mach B;
DR WPI; 90-348429/46.
DR N-PSDB; Q06469.
PT Purified proteins and compsns. - regulate expression of MHC class II
PT genes and bind to controlling DNA sequences.
PS Claim 6; Page 40; 83pp; English.
CC The sequence encodes a protein which regulates the expression of
CC MHC class II genes by binding to DNA sequences which control this
CC expression. The protein causes down-regulation, useful for the
CC prevention and treatment of autoimmune diseases such as Insulin
CC Dependent diabetes, Multiple Sclerosis, Lupus Erythematosis and
CC Rheumatoid Arthriris. The protein can also be used for screening
CC and identifying substances capable of inhibiting the expression
CC of the MHC II genes.
CC See also R08338, R08390 and R08391.
SQ Sequence 828 AA;

Query Match 2.6%; Score 7; DB 1; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGS 54
|||||||
Db 245 GGGGSGS 251

RESULT 7
R08338
ID R08338 standard; protein; 979 AA.
AC R08338;
DT 25-FEB-1991 (first entry)
DE Complete form of human RF-X protein.
KW MHC class II; down regulation; autoimmune disease.
OS Homo sapiens.

```
PN   WO9012812-A.
PD   01-NOV-1990.
PF   18-APR-1990; E00625.
PR   18-APR-1989; EP-106944.
PR   14-AUG-1989; EP-115008.
PA   (MACH/) Mach B.
PI   Mach B;
DR   WPI; 90-348429/46.
DR   N-PSDB; Q06470.
PT   Purified proteins and compsns. - regulate expression of MHC class II
PT   genes and bind to controlling DNA sequences.
PS   Claim 6; Fig 10; 83pp; English.
CC   The protein which regulates the expression of MHC class II genes by
CC   binding to DNA sequences which control this expression.  It causes
CC   down-regulation, useful for the prevention and treatment of auto-
CC   immune diseases such as Insulin Dependent diabetes, Multiple
CC   Sclerosis, Lupus Erythematosis and Rheumatoid Arthriris.  It can
CC   also be used for screening and identifying substances capable of
CC   inhibiting the expression of the MHC II genes.  It has a calculated
CC   mol. wt of 104727.
CC   See also R07661, R08390 and R08391.
SQ   Sequence   979 AA;


  Query Match             2.6%;  Score 7;  DB 1;  Length 979;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGS 54
           |||||||
Db     396 GGGGSGS 402


RESULT   8
R08390
ID   R08390 standard; protein; 437 AA.
AC   R08390;
DT   25-FEB-1991  (first entry)
DE   Truncated form of human FR-X protein.
KW   MHC class II; down regulation; autoimmune disease; HLA promoter.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          255. .376
FT                   /label=DNA binding domain
PN   WO9012812-A.
PD   01-NOV-1990.
PF   18-APR-1990; E00625.
PR   18-APR-1989; EP-106944.
PR   14-AUG-1989; EP-115008.
PA   (MACH/) Mach B.
PI   Mach B;
DR   WPI; 90-348429/46.
PT   Purified proteins and compsns. - regulate expression of MHC class II
PT   genes and bind to controlling DNA sequences.
PS   Claim 6; Fig 10; 83pp; English.
CC   The protein which regulates the expression of MHC class II genes by
CC   binding to DNA sequences which control this expression.  It causes
CC   down-regulation, useful for the prevention and treatment of auto-
CC   immune diseases such as Insulin Dependent diabetes, Multiple
CC   Sclerosis, Lupus Erythematosis and Rheumatoid Arthriris.  It can
CC   also be used for screening and identifying substances capable of
CC   inhibiting the expression of the MHC II genes.  A polypeptide
CC   comprising only the DNA binding domain (see feature table) will
CC   bind to the X box of the HLA class II promoter.
CC   See also R08338, R07661 and R08391.
SQ   Sequence   437 AA;


  Query Match             2.6%;  Score 7;  DB 1;  Length 437;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGS 54
           |||||||
Db     245 GGGGSGS 251


RESULT   9
R14702
ID   R14702 standard; Protein; 236 AA.
AC   R14702;
DT   01-FEB-1992  (first entry)
DE   Vh-Lab-Vl construction (9), single chain antibody.
KW   SCA; SCAPA; trombus; t-PA; urokinase.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   region          1. .113
FT                   /label= Vh
FT                   /note= "from MAb MA-15C5"
FT   region          114. .128
FT                   /label= Lab
FT   region          129. .236
FT                   /label= Vl
FT                   /note= "from MAb MA-15C5"
FT   region          108. .111
FT                   /label= Vh_anchor_region
FT   region          118. .122
FT                   /label= Vl_anchor_region
PN   WO9116353-A.
PD   31-OCT-1991.
PF   21-APR-1991; E00767.
PR   23-APR-1990; EP-401090.
PA   (CORV-) CORVAS INT NV.
PI   Laroche Y, Holvoet P, Demaeyer M;
DR   WPI; 91-339763/46.
PT   New single-chain MAbs specific to esp. fibrin - used for imaging,
PT   and when connected to plasminogen activating portion used for
PT   lysis of thrombi
PS   Disclosure; Fig 6; 78pp; English.
CC   The SCA can be labelled and used for the imaging of thrombi.
CC   The SCA may be connected, via a 2nd linker (Lcd) to a plasminogen
CC   activating portion (PA-portion), pref. a catalytic domain of t-PA or
CC   urokinase, esp. scuPA, to form a thrombolytic agent (SCAPA).
CC   The SCAPA has an increased half-life and has a lower mol.wt. than
CC   similar agents and is therefore expected to have a reduced
CC   immunogenicity and improved thrombus penetration.
CC   See also Q14468-70 and R14694-703 and R15690-93.
SQ   Sequence   236 AA;


  Query Match             2.6%;  Score 7;  DB 1;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db     114 GGGSGSG 120


RESULT  10
R20227
ID   R20227 standard; Protein; 223 AA.
AC   R20227;
DT   22-APR-1992  (first entry)
DE   pY53.3 (NCIMB 40308).
KW   TDF gene; Y-chromosome; mt-box; DNA-binding protein;
KW   sex determination; mating type; testis determining factor.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   binding_site    77. .156
FT                   /label= mt-protein
PN   WO9200375-A.
PD   09-JAN-1992.
PF   28-JUN-1991; G01057.
PR   28-JUN-1990; GB-014446.
PR   13-JUL-1990; GB-015488.
```

PR 09-MAY-1991; GB-010085.
PA (IMCR) IMP CANCER RES TECH.
PI Goodfellow PN, Lovell-Badge R;
DR WPI; 92-041557/05.
DR N-PSDB; Q20684.
PT Nucleic acid pY53.3, its fragments and oligo-nucleotide(s) - is
PT used for determining or controlling sex of e.g. embryos by
PT detecting Y chromosomes
PS Claim 24; Page 125 and Fig 19(1-3); 183pp; English.
CC The TDF gene comprising a Y-chromosome specific sequence contg.
CC the mating type box (mt-box). The mt-box encodes a functional testis
CC determining factor (mt-protein, see features) which when expressed at
CC the appropriate stage of embryo development results in testis
CC formation and subsequence growth of the embryo as male.
CC The sequence of the mt-protein is similar to the DNA-binding motif
CC of known DNA-binding proteins. There are a number of residues
CC conserved between human, rabbit and mouse mt-proteins, but not
CC conserved in DNA-binding proteins not associated with sex
CC determination at least at the stage of testis formation.
CC See also Q20680-81, Q20684-87, Q22530-31, Q22535-36 and Q22560-67.
SQ Sequence 223 AA;

Query Match 2.6%; Score 7; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 187 VALSLFL 193
|||||||
Db 12 VALSLFL 18

RESULT 11
R24048
ID R24048 standard; Protein; 461 AA.
AC R24048;
DT 08-DEC-1992 (first entry)
DE Heavy chain of 4D5 Fab IgG antibody - recognises HER-2 receptor.
KW humanised IgG antibody; human growth hormone; hGH; selection;
KW screening.
OS Chimaeric Homo sapiens.
PN WO9209690-A.
PD 11-JUN-1992.
PF 03-DEC-1991; U09133.
PR 03-DEC-1990; US-621667.
PR 10-APR-1991; US-683400.
PR 14-JUN-1991; US-715300.
PR 08-AUG-1991; US-743614.
PA (GETH) GENENTECH INC.
PI Bass S, Garrard LJ, Greene R, Henner DJ, Lowman HB;
PI Matthews DJ, Wells JA;
DR WPI; 92-217069/26.
DR N-PSDB; Q25592.
PT Selecting and enriching variant proteins - comprises fusing gene
PT encoding e.g. growth hormone to part of M13 phage coat protein
PT and mutagenising fusion prior to selection
PS Claim 46; Fig 11; 102pp; English.
CC This sequence represents the heavy chain Fab portion of the 4D5
CC humanised IgG antibody that recognises the HER-2 receptor. The 4D5
CC gene was inserted into the plasmid pSO132, which had the DNA encoding
CC human growth hormone excised from it. The plasmid was used to transform
CC E. coli SR101. - See Q25592.
CC Recombinant antibody was then used in a RIA for the HER-2 antigen
CC (ECD).
SQ Sequence 461 AA;

Query Match 2.6%; Score 7; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GGGSGSG 55
|||||||
Db 305 GGGSGSG 311

RESULT 12
R27097
ID R27097 standard; Protein; 484 AA.
AC R27097;
DT 02-MAR-1993 (first entry)
DE Translation of the signal:BPTI:mature III construct.
KW Bovine pancreatic trypsin inhibitor; M13; major coat protein;
KW tripartite gene; surface; gene III.
OS Synthetic.
FH Key Location/Qualifiers
FT peptide 1. .18
FT /note= "signal peptide"
FT region 19. .76
FT /note= "BPTI insertion"
FT protein 77. .456
FT /note= "mature gene III protein"
FT protein 457. .479
FT /note= "uncharged anchor region"
PN WO9215679-A.
PD 17-SEP-1992.
PF 28-FEB-1992; U01539.
PR 01-MAR-1991; US-664989.
PR 17-JUN-1991; US-715834.
PA (PROT-) PROTEIN ENG CORP.
PI Kent RB, Ladner RC, Ley AC, Roberts BL, Markland W;
PI Guterman SK;
DR WPI; 92-331725/40.
DR N-PDSB; Q28836.
PT Developing new epitope(s) or binding proteins - using a display
PT phage library which is less biased and in which irreversible
PT binding to target is alleviated
PS Example 2; Page 124; 168pp; English.
CC The protein sequence is that translated from a construct for a
CC tripartite gene comprising the M13 major coat protein III signal
CC sequence, the bovine pancreatic trypsin inhibitor gene, and the
CC mature M13 major coat protein III gene. The construct was prepd.
CC from 16 synthetic oligomers. Expression of the construct allows
CC expression of BPTI on the surface of M13 and alleviates problems
CC of irreversible binding of the display phage for the target and
CC also provides a less biased and hence more efficient phage library
CC to isolate high affinity epitopes or binding proteins.
CC See also R27083-96.
SQ Sequence 484 AA;

Query Match 2.6%; Score 7; DB 1; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GGGSGSG 55
|||||||
Db 328 GGGSGSG 334

RESULT 13
R35081
ID R35081 standard; Protein; 3080 AA.
AC R35081;
DT 28-MAY-1993 (first entry)
DE ZYMV polyprotein.
KW Zuchini yellow mosiac virus; ZYMV; potyvirus; polyprotein; protease;
KW proteolytic activity; 49 kD protease; trypsin-like cysteine protease;
KW animal picornavirus; sissile bond; NIb; protein; coat.
OS Zuchini yellow mosaic virus.
FH Key Location/Qualifiers
FT cleavage_site 766. .767
FT /note= "Cleavage site between the aphid transmission
FT helper component (HC) and the 46kD protein"
FT cleavage_site 1164. .1165
FT /note= "Cleavage site between 46 kD protein and the

```
FT                   cytoplamsic inclusion protein (CI)"
FT   cleavage_site   1798. .1799
FT                   /note= "Cleavage site between CI and VPg/protease (VPg
FT                   and protease are probably not separated in
FT                   ZYMV)"
FT   cleavage_site   2284. .2285
FT                   /note= "Cleavage site between VPg/protease and RNA
FT                   replicase (REP)"
FT   cleavage_site   2801. .2802
FT                   /note= "Cleavage site between REP and the coat
FT                   protein (CP)"
PN   WO9301305-A.
PD   21-JAN-1993.
PF   09-JUL-1992; U05745.
PR   09-JUL-1991; US-727837.
PA   (BALI/) BALINT R.
PI   Balint R;
DR   WPI; 93-045506/05.
DR   N-PSDB; Q35297.
PT   Method for identifying protease inhibitors - useful for drugs
PT   screening for treating e.g. chronic inflammation, metastatic
PT   cancers and viral infections
PS   Disclosure;; 62pp; English.
CC   This sequence was not disclosed in the specification but was decoded
CC   from Q35297.  This sequence represents the zuchini yellow mosiac virus
CC   (ZYMV) polyprotein.  ZYMV is a potyvirus and expresses its genome as a
CC   single 350 kD polyprotein which is cleaved into at least seven mature
CC   gene products by three distinct proteolytic activities.  Two of the
CC   proteases are virus encoded, including the potyviral 49 kD protease.
CC   This protease is responsible for at least five of the seven cleavages.
CC   This enzyme is a trypsin-like cysteine protease which is structurally
CC   and mechanistically representative of the largest class of viral
CC   proteases, including those of the animal picornaviruses.  This enzyme
CC   is highly specific and appears to recognise a region comprised of
CC   about seven amino acids surrounding the sissile bond.  Of the five
CC   sites cleaved by this enzyme, the two flanking the protease appear to
CC   be cleaved intramolecularly, while the remaining three appear to be
CC   cleaved intermolcularly.  Of the latter three, the site between the
CC   NIb protein and the coat protein appears to be the most active.
SQ   Sequence   3080 AA;


  Query Match             2.6%;  Score 7;  DB 1;  Length 3080;
  Best Local Similarity   100.0%;  Pred. No. 6.9e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 GSGSGEK 57
           |||||||
Db    2830 GSGSGEK 2836


RESULT  14
R50174
ID   R50174 standard; Protein; 211 AA.
AC   R50174;
DT   04-OCT-1994  (first entry)
DE   Membrane anchor peptide derived from M13 coat protein (gene III).
KW   Expression; insertion sequence; Vh chain; variable heavy chain;
KW   secretion signal; epitope tag; expression vector; lamda Zap;
KW   lambda Hc2; membrane; anchor peptide; bacteriophage; phage; M13;
KW   cpIII.
OS   Bacteriophage M13.
PN   WO9405781-A.
PD   17-MAR-1994.
PF   03-SEP-1993; U08364.
PR   04-SEP-1992; US-941369.
PA   (SCRI ) SCRIPPS RES INST.
PI   Lerner RA,  Light JP;
DR   WPI; 94-101186/12.
DR   N-PSDB; Q44655.
PT   Filamentous phage comprising a heterologous polypeptide and a
PT   hetero:dimer - is used to detect the presence of a preselected
PT   ligand in a sample
PS   Disclosure; Page 176-177; 232pp; English.
CC   A filamentous phage constructed so as to be able to detect the
CC   presence of a preselected ligand in a sample comprises (1) a
CC   heterologous polypeptide fused to a first filamentous phage coat
CC   protein membrane anchor and (2) a heterodimeric receptor
CC   comprising first and second receptor polypeptides, one of which
CC   is fused to a second filamentous phage coat protein membrane
CC   anchor.  This sequence is a membrane anchor polypeptide derived
CC   from the M13 coat protein gene 3 (cpIII).
SQ   Sequence   211 AA;


  Query Match             2.6%;  Score 7;  DB 1;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db      55 GGGSGSG 61


RESULT  15
R54281
ID   R54281 standard; Protein; 211 AA.
AC   R54281;
DT   10-NOV-1994  (first entry)
DE   M13 phage coat protein cpIII membrane anchor domain.
KW   Human immunodeficiency virus; HIV1; glycoprotein gp41; epitope;
KW   neutralisation; monoclonal antibody; light chain; variable region;
KW   filamentous phage M13; M13mp18; coat protein cpIII.
OS   Bacteriophage M13.
FH   Key             Location/Qualifiers
FT   misc_difference 183
FT                   /note= "no corresp. codon in Q64222"
PN   WO9407922-A.
PD   14-APR-1994.
PF   30-SEP-1993; U09328.
PR   30-SEP-1992; US-954148.
PA   (SCRI ) SCRIPPS RES INST.
PI   Barbas CF,  Burton DR,  Lerner RA;
DR   WPI; 94-135516/16.
DR   N-PSDB; Q64222.
PT   New human monoclonal antibodies neutralising HIV - react with
PT   gp120 or gp41 and nucleic acid encoding them, useful for in vivo
PT   or in vitro diagnosis and for passive immuno-therapy
PS   Example 1; Page 145-146; 248pp; English.
CC   M13mp18 replicative form DNA was used as a template for isolating
CC   the gene encoding the membrane anchor domain at cpIII. Two PCR
CC   amplifications were performed for construction of a DNA fragment
CC   consisting of the mature gene for cpIII membrane anchor domain
CC   located 5' to a sequence encoding the lacZ promoter, operator and
CC   cap-binding site for controlling light chain expression.
SQ   Sequence   211 AA;


  Query Match             2.6%;  Score 7;  DB 1;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      49 GGGSGSG 55
           |||||||
Db      55 GGGSGSG 61


Search completed: February 11, 2000, 05:14:32
Job time: 730 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 13:27:02 ; Search time 506.51 Seconds
(without alignments)
6037.961 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: OLIGO_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

d size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	550	67.9	574	50	AI680904	AI680904 tx42f05.x
2	505	62.3	556	47	AI479764	AI479764 tm69b04.x
3	493	60.9	599	62	AI923178	AI923178 wn67b10.x
4	479	59.1	537	48	AI580361	AI580361 tm47h02.x
5	474	58.5	548	38	AA772225	AA772225 ai41c01.s
6	456	56.3	507	50	AI682204	AI682204 wa71b06.x
7	447	55.2	498	45	AI349520	AI349520 qp72g05.x

8	441	54.4	530	50	AI674462	AI674462 wc44e01.x
9	439	54.2	439	44	AI299154	AI299154 qm96f11.x
10	436	53.8	471	43	AI168073	AI168073 oz89c02.x
11	423	52.2	423	44	AI299329	AI299329 qn22c06.x
12	395	48.8	395	44	AI245878	AI245878 qk27f08.x
13	389	48.0	448	41	AI038331	AI038331 ox84a06.x
14	368	45.4	500	50	AI680969	AI680969 tx35h07.x
15	320	39.5	322	62	AI903935	AI903935 MR-BT038-
16	320	39.5	398	64	AW071521	AW071521 xa36a07.x
17	319	39.4	323	62	AI903938	AI903938 MR-BT038-
18	319	39.4	323	62	AI903940	AI903940 MR-BT038-
19	311	38.4	507	25	N47594	N47594 yy59a07.s1
c 20	306	37.8	632	41	AI057115	AI057115 oz23g10.x
21	301	37.2	335	43	AI194064	AI194064 qe78d08.x
22	281	34.7	461	50	AI693751	AI693751 wd44h04.x
23	280	34.6	392	38	AA757694	AA757694 zg38h03.s
24	277	34.2	382	32	AA345598	AA345598 EST51648
25	267	33.0	385	25	D81362	D81362 HUM159G08B
26	257	31.7	300	62	AI887092	AI887092 wl96f10.x
27	250	30.9	508	36	AA639448	AA639448 nq88g05.s
28	242	29.9	252	24	N36461	N36461 yx83e05.r1
29	226	27.9	540	44	AI284315	AI284315 qj65c02.x
30	214	26.4	464	24	N34957	N34957 yy50b11.s1
c 31	202	24.9	590	69	AW140060	AW140060 UI-H-BI1-
32	201	24.8	580	23	R99199	R99199 yq64d10.r1
33	185	22.8	674	29	AA143062	AA143062 zl49h02.s
34	182	22.5	416	38	AA757905	AA757905 zg42h09.s
35	172	21.2	474	43	AI239974	AI239974 qh44d02.x
c 36	169	20.9	305	40	AA927408	AA927408 om27d10.s
37	165	20.4	216	30	AA253249	AA253249 zr53a03.r
38	165	20.4	565	46	AI420901	AI420901 tf03g10.x
39	160	19.8	475	31	AA306979	AA306979 EST178050
c 40	160	19.8	482	51	AI732252	AI732252 ye35c09.x
41	148	18.3	346	22	R26459	R26459 yh50e12.s1
42	140	17.3	480	24	N28591	N28591 yx38c03.r1
43	129	15.9	351	31	AA302858	AA302858 EST12947
44	122	15.1	417	23	D61247	D61247 HUM172D06B
c 45	113	14.0	450	41	AI066635	AI066635 oz82c10.x

ALIGNMENTS

RESULT 1
AI680904
LOCUS AI680904 574 bp mRNA EST 26-MAY-1999
DEFINITION tx42f05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272257 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION AI680904
VERSION AI680904.1 GI:4891086
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 574)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1134035.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
FEATURES Location/Qualifiers
source 1. .574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2272257"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 148 a 118 c 136 g 172 t
ORIGIN

Query Match 67.9%; Score 550; DB 50; Length 574;
Best Local Similarity 100.0%; Pred. No. 1.7e-288;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    235 GCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 GCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTT 84

Qy    295 GCCACCTCCGCCGGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATAT 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GCCACCTCCGCCGGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATAT 144

Qy    355 ATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTAC 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 ATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTAC 204

Qy    415 ACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAAT 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 ACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAAT 264

Qy    475 GAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTA 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 GAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTA 324

Qy    535 AATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 AATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCA 384

Qy    595 GATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTT 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 GATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTT 444

Qy    655 TGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCA 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 TGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCA 504

Qy    715 GATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACT 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 GATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACT 564

Qy    775 AATGAAACAT 784
          ||||||||||
Db    565 AATGAAACAT 574
```

RESULT 2
AI479764
LOCUS AI479764 556 bp mRNA EST 14-APR-1999
DEFINITION tm69b04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163343 3', mRNA sequence.

```
ACCESSION   AI479764
VERSION     AI479764.1  GI:4372932
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188819.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 614   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2163343"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      148 a    109 c    130 g    168 t      1 others
ORIGIN
```

```
  Query Match             62.3%;  Score 505;  DB 47;  Length 556;
  Best Local Similarity   99.8%;  Pred. No. 5.4e-264;
  Matches  555;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    234 GGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGT 293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGT 60

Qy    294 TGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATA 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATA 120

Qy    354 TATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTA 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTA 180

Qy    414 CACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAA 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAA 240

Qy    474 TGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGT 533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGT 300

Qy    534 AAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGC 593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGC 360

Qy    594 AGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTT 653
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGATCGATTNTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTT 420

Qy    654 TTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTC 713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTC 480

Qy    714 AGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTAC 773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTAC 540

Qy    774 TAATGAAACATTTAGA 789
          ||||||||||||||||
Db    541 TAATGAAACATTTAGA 556
```

```
RESULT   3
AI923178
LOCUS       AI923178      599 bp    mRNA            EST       02-SEP-1999
DEFINITION  wn67b10.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2450491 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI923178
VERSION     AI923178.1  GI:5659142
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133491.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2450491"
                     /clone_lib="NCI_CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I  and Eco RI sites of the modified
                     pT7T3 vector.  Library went througb one round of
```

normalization. Library constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 148 a 127 c 146 g 176 t 2 others
ORIGIN

Query Match 60.9%; Score 493; DB 62; Length 599;
Best Local Similarity 99.7%; Pred. No. 1.8e-257;
Matches 593; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCCACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    421 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTANGGTTTTGTGGAATTGGGAGC 480

Qy    670 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTAC 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTAC 540

Qy    730 ATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACAT 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACAT 595
```

RESULT 4
AI580361
LOCUS AI580361 537 bp mRNA EST 05-APR-1999
DEFINITION tm47h02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2161299 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION AI580361
VERSION AI580361.1 GI:4564737
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 537)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949103.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 492.
FEATURES Location/Qualifiers
source 1. .537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2161299"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 124 a 117 c 139 g 156 t 1 others
ORIGIN

Query Match 59.1%; Score 479; DB 48; Length 537;
Best Local Similarity 99.8%; Pred. No. 7.6e-250;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGATACCCTGCTNTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 480

Qy    670 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGG 719
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGG 530
```

RESULT 5

```
AA772225
LOCUS       AA772225      548 bp    mRNA            EST       31-DEC-1998
DEFINITION  ai41c01.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            1359552 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;,
            mRNA sequence.
ACCESSION   AA772225
VERSION     AA772225.1  GI:2824008
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811076.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Insert Length: 978    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="1359552"
                     /clone_lib="Soares_parathyroid_tumor_NbHPA"
                     /tissue_type="parathyroid tumor"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: parathyroid gland; Vector: pT7T3D
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer
                     [5'-
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTTT
                     T-3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                     adenomas was kindly provided by Dr. Stephen Marx, National
                     Institute of Diabetes and Digestive and Kidney Diseases,
                     NIH."
BASE COUNT      146 a    106 c    126 g    169 t      1 others
ORIGIN


  Query Match             58.5%;  Score 474;  DB 38;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 4e-247;
  Matches  474;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    315 GGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAAT 374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 GGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAAT 134

Qy    375 AAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTT 434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 AAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTT 194

Qy    435 TCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAA 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 TCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAA 254

Qy    495 CGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGT 554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 CGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGT 314

Qy    555 GGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATA 614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 GGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATA 374

Qy    615 CCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAAT 674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    375 CCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAAT 434

Qy    675 TGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTAT 734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 TGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTAT 494

Qy    735 AGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACATTTAG 788
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 AGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACATTTAG 548


RESULT   6
AI682204
LOCUS       AI682204      507 bp    mRNA            EST       26-MAY-1999
DEFINITION  wa71b06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2301587 3', mRNA sequence.
ACCESSION   AI682204
VERSION     AI682204.1  GI:4892386
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1136077.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2301587"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      120 a    109 c    125 g    152 t      1 others
ORIGIN
```

```
  Query Match             56.3%;  Score 456;  DB 50;  Length 507;
  Best Local Similarity   99.8%;  Pred. No. 2.5e-237;
  Matches  506;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    208 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 60

Qy    268 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 120

Qy    328 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 180

Qy    388 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 240

Qy    448 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

Qy    508 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 360

Qy    568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 420

Qy    628 TTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 687
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    421 TTGTTAAAGTTTTGCACTGTANGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 480

Qy    688 ATTTCAATGCAGATTGTTGGACCTTCA 714
          |||||||||||||||||||||||||||
Db    481 ATTTCAATGCAGATTGTTGGACCTTCA 507


RESULT   7
AI349520
LOCUS       AI349520      498 bp    mRNA            EST       03-FEB-1999
DEFINITION  qp72g05.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1928600 3', mRNA sequence.
ACCESSION   AI349520
VERSION     AI349520.1  GI:4086726
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1009   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1928600"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT      113 a    112 c    128 g    144 t      1 others
ORIGIN

  Query Match             55.2%;  Score 447;  DB 45;  Length 498;
  Best Local Similarity   99.8%;  Pred. No. 2e-232;
  Matches  497;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGATACCCTGCTNTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 480

Qy    670 CTAATTGATTTCATTCTT 687
          ||||||||||||||||||
Db    481 CTAATTGATTTCATTCTT 498


RESULT   8
AI674462
LOCUS       AI674462      530 bp    mRNA            EST       19-MAY-1999
DEFINITION  wc44e01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321496 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI674462
VERSION     AI674462.1  GI:4874942
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

```
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948555.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2321496"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      123 a    116 c    133 g    157 t      1 others
ORIGIN


  Query Match             54.4%;  Score 441;  DB 50;  Length 530;
  Best Local Similarity   99.8%;  Pred. No. 3.7e-229;
  Matches  491;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    232 ACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCT 291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39 ACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCT 98

Qy    292 GTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAA 351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 GTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAA 158

Qy    352 TATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAAC 411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 TATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAAC 218

Qy    412 TACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGC 471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 TACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGC 278

Qy    472 AATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAAT 531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 AATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAAT 338

Qy    532 GTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGA 591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 GTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGA 398

Qy    592 GCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGG 651
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    399 GCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGNTTTGCACTGTAGGG 458

Qy    652 TTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCT 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 TTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCT 518

Qy    712 TCAGATGGAAGT 723
          ||||||||||||
Db    519 TCAGATGGAAGT 530


RESULT   9
AI299154
LOCUS       AI299154      439 bp    mRNA            EST       29-JAN-1999
DEFINITION  qm96f11.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1896621 3',
            mRNA sequence.
ACCESSION   AI299154
VERSION     AI299154.1  GI:3958808
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 966   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1896621"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      105 a    100 c    110 g    124 t
ORIGIN


  Query Match             54.2%;  Score 439;  DB 44;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-228;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    208 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 60

Qy    268 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 120

Qy    328 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 180

Qy    388 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 240

Qy    448 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

Qy    508 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 360

Qy    568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 420

      628 TTGTTAAAGTTTTGCACTG 646
          ||||||||||||||||||
Db    421 TTGTTAAAGTTTTGCACTG 439


RESULT 10
AI168073
LOCUS       AI168073      471 bp    mRNA            EST       05-OCT-1998
DEFINITION  oz89c02.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
            clone IMAGE:1682498 3', mRNA sequence.
ACCESSION   AI168073
VERSION     AI168073.1  GI:3701243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2282128.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1682498"
                     /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      111 a    110 c    120 g    130 t
ORIGIN


  Query Match           53.8%;  Score 436;  DB 43;  Length 471;
  Best Local Similarity  100.0%;  Pred. No. 2e-226;
  Matches  436;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGG 625
          ||||||||||||||||
Db    421 GGATACCCTGCTTTGG 436


RESULT 11
AI299329
LOCUS       AI299329      423 bp    mRNA            EST       01-FEB-1999
DEFINITION  qn22c06.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1898986 3',
            mRNA sequence.
ACCESSION   AI299329
VERSION     AI299329.1  GI:3958983
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1021   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1898986"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
```

/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 101 a 98 c 107 g 117 t
ORIGIN

Query Match 52.2%; Score 423; DB 44; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.4e-219;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    208 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 60

Qy    268 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 120

Qy    328 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 180

Qy    388 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 240

Qy    448 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

Qy    508 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 360

Qy    568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 420

Qy    628 TTG 630
          |||
Db    421 TTG 423
```

RESULT 12
AI245878
LOCUS AI245878 395 bp mRNA EST 28-JAN-1999
DEFINITION qk27f08.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1870215 3', mRNA sequence.
ACCESSION AI245878
VERSION AI245878.1 GI:3841275
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 395)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152532.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1000 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 389.
FEATURES Location/Qualifiers
source 1. .395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1870215"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 94 a 98 c 100 g 103 t
ORIGIN

Query Match 48.8%; Score 395; DB 44; Length 395;
Best Local Similarity 100.0%; Pred. No. 4.2e-204;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATG 584
          |||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATG 395
```

RESULT 13
AI038331
LOCUS AI038331 448 bp mRNA EST 24-SEP-1998
DEFINITION ox84a06.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:1662994 3', mRNA sequence.
ACCESSION AI038331
VERSION AI038331.1 GI:3277525
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:802984.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1099   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 432.
FEATURES              Location/Qualifiers
     source           1. .448
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1662994"
                      /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                      /tissue_type="senescent fibroblast"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                      RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT      118 a     92 c    105 g    133 t
ORIGIN


  Query Match             48.0%;  Score 389;  DB 41;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-201;
  Matches  389;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    260 TCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGT 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 TCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGT 62

Qy    320 CGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATG 379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 CGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATG 122

Qy    380 ACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 ACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 182

Qy    440 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAG 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAG 242

Qy    500 TTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAG 559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 TTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAG 302

Qy    560 TCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTG 619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTG 362

Qy    620 CTTTGGGTTTGTTAAAGTTTTGCACTGTA 648
          |||||||||||||||||||||||||||||
Db    363 CTTTGGGTTTGTTAAAGTTTTGCACTGTA 391


RESULT  14
```

```
AI680969
LOCUS       AI680969      500 bp    mRNA            EST       26-MAY-1999
DEFINITION  tx35h07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2271613 3',
            mRNA sequence.
ACCESSION   AI680969
VERSION     AI680969.1  GI:4891151
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135321.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 500.
FEATURES              Location/Qualifiers
     source           1. .500
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2271613"
                      /clone_lib="NCI_CGAP_Lu24"
                      /tissue_type="carcinoid"
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Plasmid DNA from the normalized
                      library NCI_CGAP_Lu5 was prepared, and ss circles were
                      made in vitro. Following HAP purification, this DNA was
                      used as tracer in a subtractive hybridization reaction.
                      The driver was PCR-amplified cDNAs from a pool of 5,000
                      clones made from the same library (cloneIDs
                      1414920-1417991 and 1520904-1522439). Subtraction by Bento
                      Soares and M. Fatima Bonaldo.  "
BASE COUNT      114 a    112 c    125 g    145 t      4 others
ORIGIN


  Query Match             45.4%;  Score 368;  DB 50;  Length 500;
  Best Local Similarity   99.8%;  Pred. No. 2e-189;
  Matches  418;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    230 TGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 TGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

Qy    290 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGAC 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGAC 160

Qy    350 AATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAA 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 AATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAA 220

Qy    410 ACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTG 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 ACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTG 280

Qy    470 GCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAA 529
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 GCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAA 340

Qy    530 ATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 ATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGN 400

Qy    590 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTA 648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTA 459


RESULT  15
AI903935
LOCUS       AI903935      322 bp    mRNA            EST       01-DEC-1999
DEFINITION  MR-BT038-080199-049 BT038 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AI903935
VERSION     AI903935.1  GI:6494231
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1134501.
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/seq/gethtml.pl?t1=MR&t2=MR-BT038-
            049.html&t3=080199&t4=1)
            Seq primer: puc 18 forward.
FEATURES             Location/Qualifiers
     source          1. .322
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT038"
                     /sex="female"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No.
                     196,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       90 a     60 c     68 g    104 t
ORIGIN


  Query Match           39.5%;  Score 320;  DB 62;  Length 322;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-163;
  Matches  320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    343 GTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAAC 402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 GTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAAC 62

Qy    403 TGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGAT 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 TGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGAT 122

Qy    463 TCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCT 522
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 TCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCT 182

Qy    523 TGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGA 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 TGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGA 242

Qy    583 TGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGC 642
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 TGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGC 302

Qy    643 ACTGTAGGGTTTTGTGGAAT 662
          ||||||||||||||||||||
Db    303 ACTGTAGGGTTTTGTGGAAT 322


Search completed: February 10, 2000, 14:59:09
Job time: 5527 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 10, 2000, 17:01:07 ; Search time 11.6 Seconds
(without alignments)
297.955 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 1439
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 130275 seqs, 12848600 residues

base : Issued_Patents_AA:*

Word size : 0

Number of hits that pass the threshold : 130275
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	85	5.9	338	2	US-08-414-657D-42	Sequence 42, Appl
2	85	5.9	338	2	US-08-414-657D-43	Sequence 43, Appl
3	85	5.9	310	2	US-08-414-657D-45	Sequence 45, Appl
4	83	5.8	764	2	US-08-177-109A-2	Sequence 2, Appli
5	83	5.8	764	2	US-08-687-706-2	Sequence 2, Appli
6	80.5	5.6	1237	1	US-08-241-853-2	Sequence 2, Appli
7	80.5	5.6	1237	2	US-08-850-917-2	Sequence 2, Appli
8	80.5	5.6	325	2	US-08-414-657D-2	Sequence 2, Appli
9	80.5	5.6	325	2	US-08-414-657D-41	Sequence 41, Appl
10	80.5	5.6	304	2	US-08-414-657D-44	Sequence 44, Appl
11	79.5	5.5	354	2	US-08-700-013B-9	Sequence 9, Appli
12	79.5	5.5	797	2	US-08-700-013B-21	Sequence 21, Appl
13	79	5.5	797	2	US-08-700-013B-19	Sequence 19, Appl
14	78.5	5.5	338	2	US-08-414-657D-60	Sequence 60, Appl
15	76	5.3	525	1	US-08-356-340-2	Sequence 2, Appli
16	76	5.3	525	2	US-08-786-555-2	Sequence 2, Appli
17	74	5.1	2973	2	US-08-821-355A-7	Sequence 7, Appli
18	74	5.1	1193	2	US-08-400-159-10	Sequence 10, Appl
19	74	5.1	344	2	US-07-857-224B-93	Sequence 93, Appl
20	73.5	5.1	3033	1	US-07-925-695-8	Sequence 8, Appli
21	73.5	5.1	191	2	US-08-290-665A-177	Sequence 177, App
22	73.5	5.1	191	3	PCT-US95-10398-177	Sequence 177, App
23	73	5.1	447	1	US-08-188-281B-20	Sequence 20, Appl
24	73	5.1	381	1	US-08-173-508-12	Sequence 12, Appl
25	73	5.1	627	1	US-08-295-814E-4	Sequence 4, Appli
26	73	5.1	627	1	US-08-291-299-10	Sequence 10, Appl
27	73	5.1	381	2	US-08-265-310-12	Sequence 12, Appl
28	73	5.1	627	3	PCT-US93-01959-4	Sequence 4, Appli
29	73	5.1	447	3	PCT-US94-07280-20	Sequence 20, Appl
30	73	5.1	447	3	PCT-US95-01087-20	Sequence 20, Appl
31	73	5.1	627	3	PCT-US95-10579-10	Sequence 10, Appl
32	73	5.1	328	4	5212074-4	Patent No. 5212074
33	72.5	5.0	3033	1	US-07-925-695-9	Sequence 9, Appli
34	72.5	5.0	615	2	US-08-752-307B-9	Sequence 9, Appli
35	72	5.0	299	2	US-08-872-437-2	Sequence 2, Appli
36	71.5	5.0	602	1	US-08-295-814E-2	Sequence 2, Appli
37	71.5	5.0	602	1	US-08-291-299-9	Sequence 9, Appli
38	71.5	5.0	602	3	PCT-US93-01959-2	Sequence 2, Appli
39	71.5	5.0	602	3	PCT-US95-10579-9	Sequence 9, Appli
40	71	4.9	133	1	US-07-924-753-15	Sequence 15, Appl
41	71	4.9	620	1	US-08-325-071-65	Sequence 65, Appl
42	71	4.9	650	1	US-08-325-071-67	Sequence 67, Appl
43	71	4.9	2476	2	US-08-276-967-2	Sequence 2, Appli
44	71	4.9	589	2	US-08-317-305-4	Sequence 4, Appli
45	71	4.9	405	2	US-08-700-013B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-414-657D-42
; Sequence 42, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-42

Query Match 5.9%; Score 85; DB 2; Length 338;
Best Local Similarity 27.7%; Pred. No. 0.79;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

```
Qy    101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
          : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db    230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy    157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
          || |    |::|        |:| | | :|| |  :|| |     ||           |
```

```
Db    286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----------LA 328

Qy    210 LLKFCTVGFC 219
          || :  |
Db    329 ASLFCLLSKC 338


RESULT   2
US-08-414-657D-43
; Sequence 43, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-414-657D-43


  Query Match             5.9%;  Score 85;  DB 2;  Length 338;
  Best Local Similarity   27.7%;  Pred. No. 0.79;
  Matches   36;  Conservative   15;  Mismatches   47;  Indels   32;  Gaps    7;

Qy    101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
          : | : |||||   |      ::   | :|| |     :  |       |     |:| ::
Db    230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy    157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
          || |    |::|        |:| | | :|| |  :|| |     ||           |
Db    286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----------LA 328

Qy    210 LLKFCTVGFC 219
          || :  |
Db    329 ASLFCLLSKC 338


RESULT   3
```

```
US-08-414-657D-45
; Sequence 45, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  310 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-414-657D-45


  Query Match             5.9%;  Score 85;  DB 2;  Length 310;
  Best Local Similarity   27.7%;  Pred. No. 0.7;
  Matches   36;  Conservative   15;  Mismatches   47;  Indels   32;  Gaps    7;

Qy    101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
          : | : |||||   |      ::   | :|| |     :  |       |     |:| ::
Db    202 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 257

Qy    157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
          || |    |::|        |:| | | :|| |  :|| |     ||           |
Db    258 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----------LA 300

Qy    210 LLKFCTVGFC 219
          || :  |
Db    301 ASLFCLLSKC 310


RESULT   4
US-08-177-109A-2
; Sequence 2, Application US/08177109A
; Patent No. 5869615
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/177,109A
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-177-109A-2


  Query Match             5.8%;  Score 83;  DB 2;  Length 764;
  Best Local Similarity   24.1%;  Pred. No. 4.1;
  Matches   49;  Conservative   21;  Mismatches   71;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV---------- 61
           |: | |       | |:|| | || ||            || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :    |||  |   :  | |          :|| |   : |            :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


RESULT   5
US-08-687-706-2
; Sequence 2, Application US/08687706
; Patent No. 5928892
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,706
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/177,109
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107 DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-687-706-2


  Query Match             5.8%;  Score 83;  DB 2;  Length 764;
  Best Local Similarity   24.1%;  Pred. No. 4.1;
  Matches   49;  Conservative   21;  Mismatches   71;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV---------- 61
           |: | |       | |:|| | || ||            || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :    |||  |   :  | |          :|| |   : |            :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


RESULT   6
US-08-241-853-2
; Sequence 2, Application US/08241853
; Patent No. 5693488
;  GENERAL INFORMATION:
;    APPLICANT:  Fang, Kathy S.
;    APPLICANT:  Hanafusa, Hidesaburo
;    TITLE OF INVENTION:  A TRANSMEMBRANE TYROSINE PHOSPHATASE
;    TITLE OF INVENTION:  AND METHODS OF USE THEREOF
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/241,853
;      FILING DATE:  12-MAY-1994
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1237 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Gallus domesticus
;      STRAIN:  DT40
U[redacted]-241-853-2

  Query Match             5.6%;  Score 80.5;  DB 1;  Length 1237;
  Best Local Similarity   21.6%;  Pred. No. 15;
  Matches   48;  Conservative   30;  Mismatches   75;  Indels   69;  Gaps   11;

Qy      71 PSAPEAVTA----RLVGVLWFVSVTTGPWGAVATSAGG--------------EESLKCE 111
           | ||| :|     | | | |       | | :     |               | | | |
Db     323 PKAPENLTVHPTDRNVTVTWM-----KPTGTLEKHIDGYTVECNNTSQNVNRNETSFTCG 377

Qy     112 DL------------KVGQYICKD---PKINDATQ----EPVNCTNYTAHVSCFPAPNIT 151
           ||            |  :|  |:    |:| : |    :| | |::   :       :
Db     378 DLEPYSTGSVSVRAFKKSKYKNKNFEGEKVNGSFQTKPAKPENVTDFKLTLLADNTVKVA 437

Qy     152 CKDSS--GNETHF-------------TGNEVGFFKPISCRNVN---GYSYKVAVALSLF 192
           |:     |||| |                 ||  |    : |:::    |::|::|   ::
Db     438 CRSQKVYGNETKFKLSWNSSSNSGENQRKNECNF----TVRDLSYLTKYTFKISVFNGVY 493

Qy     193 LGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQI 234
           |    :  |  | :  |: |    |  : : |  :|:  :|
Db     494 TGDSVCEEIYTRYNSRALIIFLV--FLIVVTSIALLLVLYKI 533


RESULT   7
US-08-850-917-2
; Sequence 2, Application US/08850917
; Patent No. 5854045
; GENERAL INFORMATION:
;    APPLICANT:  Fang, Kathy S.
;    APPLICANT:  Hanafusa, Hidesaburo
;    TITLE OF INVENTION:  A TRANSMEMBRANE TYROSINE PHOSPHATASE
;    TITLE OF INVENTION:  AND METHODS OF USE THEREOF
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,917
;      FILING DATE:  02-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/241,853
;      FILING DATE:  12-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1237 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Gallus domesticus
;      STRAIN:  DT40
US-08-850-917-2

  Query Match             5.6%;  Score 80.5;  DB 2;  Length 1237;
  Best Local Similarity   21.6%;  Pred. No. 15;
  Matches   48;  Conservative   30;  Mismatches   75;  Indels   69;  Gaps   11;

Qy      71 PSAPEAVTA----RLVGVLWFVSVTTGPWGAVATSAGG--------------EESLKCE 111
           | ||| :|     | | | |       | | :     |               | | | |
Db     323 PKAPENLTVHPTDRNVTVTWM-----KPTGTLEKHIDGYTVECNNTSQNVNRNETSFTCG 377

Qy     112 DL------------KVGQYICKD---PKINDATQ----EPVNCTNYTAHVSCFPAPNIT 151
           ||            |  :|  |:    |:| : |    :| | |::   :       :
Db     378 DLEPYSTGSVSVRAFKKSKYKNKNFEGEKVNGSFQTKPAKPENVTDFKLTLLADNTVKVA 437

Qy     152 CKDSS--GNETHF-------------TGNEVGFFKPISCRNVN---GYSYKVAVALSLF 192
           |:     |||| |                 ||  |    : |:::    |::|::|   ::
Db     438 CRSQKVYGNETKFKLSWNSSSNSGENQRKNECNF----TVRDLSYLTKYTFKISVFNGVY 493

Qy     193 LGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQI 234
           |    :  |  | :  |: |    |  : : |  :|:  :|
Db     494 TGDSVCEEIYTRYNSRALIIFLV--FLIVVTSIALLLVLYKI 533


RESULT   8
US-08-414-657D-2
; Sequence 2, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
```

```
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-414-657D-2


  Query Match             5.6%;  Score 80.5;  DB 2;  Length 325;
  Best Local Similarity   29.9%;  Pred. No. 2.1;
  Matches   32;  Conservative   14;  Mismatches   40;  Indels   21;  Gaps    6;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db     223 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 278

Qy     157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWL 196
           || |    |::|        |:| | | :|| |  :|| |     ||
Db     279 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL 319


RESULT  9
US-08-414-657D-41
; Sequence 41, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-414-657D-41


  Query Match             5.6%;  Score 80.5;  DB 2;  Length 325;
  Best Local Similarity   29.9%;  Pred. No. 2.1;
  Matches   32;  Conservative   14;  Mismatches   40;  Indels   21;  Gaps    6;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db     223 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 278

Qy     157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWL 196
           || |    |::|        |:| | | :|| |  :|| |     ||
Db     279 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL 319


RESULT  10
US-08-414-657D-44
; Sequence 44, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  304 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-414-657D-44


  Query Match             5.6%;  Score 80.5;  DB 2;  Length 304;
  Best Local Similarity   29.9%;  Pred. No. 1.9;
  Matches   32;  Conservative   14;  Mismatches   40;  Indels   21;  Gaps    6;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db     202 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 257
```

```
Qy    157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWL 196
          || |    |::|        |:| | | :|| |  :|| |     ||
Db    258 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL 298


RESULT  11
US-08-700-013B-9
; Sequence 9, Application US/08700013B
; Patent No. 5919653
;  GENERAL INFORMATION:
;    APPLICANT:  Albert, Vivian R.
;    APPLICANT:  Kowalski, Leslie R.Z.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  McKelvy, Jeffrey F.
;    TITLE OF INVENTION:  Human Glycine Transporter
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/700,013B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  354 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-700-013B-9


  Query Match             5.5%;  Score 79.5;  DB 2;  Length 354;
  Best Local Similarity   23.0%;  Pred. No. 3;
  Matches   60;  Conservative   38;  Mismatches  110;  Indels   53;  Gaps   13;

Qy      37 LALLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLW---------- 87
           || ||    :  |   |   ||| |   |   |      : : |: : :
Db      48 LAGLPIFFLEVSLGQFASQGPVSVWKAIPAL-QGCGIAMLINSVLIAIYYNVIICYTLFY 106

Qy      87 ----FVSVTTGPWGAVATSAGGEESLKCED---LKVGQYICKD-PKINDATQEPVNCTNY 138
                ||||   |||:        |   |:|   | :   :  | |||          |
Db     107 LFASFVSVL--PWGSCNNPWNTPE---CKDKTKLLLDSCVISDHPKI--------QIKNS 153

Qy     139 TAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSY----KVAVALSLFLG 194
           |  :: :|   :    |  |:|  :|:|  :||    :   |  |    :  :|| |||
Db     154 TFCMTAYPNVTMVNFTSQANKTFVSGSE-EYFKYFVLKISAGIEYPGEIRWPLALCLFLA 212

Qy     195 WLGADRFYLGYPAL-------GLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYIIDY 247
           |:      : | :|       | : : |  |  :  |:  ::  : : |   |  | |
Db     213 WV------IVYASLAKGIKTSGKVVYFTATFPYV-VLVILLIRGVTLPGAGAGIWYFITP 265

Qy     248 YGTRLTRLSITNETFRKTQLY 268
```

```
           :||  ::  :    ||::
Db     266 KWEKLTDATVWKDA--ATQIF 284


RESULT  12
US-08-700-013B-21
; Sequence 21, Application US/08700013B
; Patent No. 5919653
;  GENERAL INFORMATION:
;    APPLICANT:  Albert, Vivian R.
;    APPLICANT:  Kowalski, Leslie R.Z.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  McKelvy, Jeffrey F.
;    TITLE OF INVENTION:  Human Glycine Transporter
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/700,013B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  797 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-700-013B-21


  Query Match             5.5%;  Score 79.5;  DB 2;  Length 797;
  Best Local Similarity   23.0%;  Pred. No. 9.7;
  Matches   60;  Conservative   38;  Mismatches  110;  Indels   53;  Gaps   13;

Qy      37 LALLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLW---------- 87
           || ||    :  |   |   ||| |   |   |      : : |: : :
Db     239 LAGLPIFFLEVSLGQFASQGPVSVWKAIPAL-QGCGIAMLINSVLIAIYYNVIICYTLFY 297

Qy      87 ----FVSVTTGPWGAVATSAGGEESLKCED---LKVGQYICKD-PKINDATQEPVNCTNY 138
                ||||   |||:        |   |:|   | :   :  | |||          |
Db     298 LFASFVSVL--PWGSCNNPWNTPE---CKDKTKLLLDSCVISDHPKI--------QIKNS 344

Qy     139 TAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSY----KVAVALSLFLG 194
           |  :: :|   :    |  |:|  :|:|  :||    :   |  |    :  :|| |||
Db     345 TFCMTAYPNVTMVNFTSQANKTFVSGSE-EYFKYFVLKISAGIEYPGEIRWPLALCLFLA 403

Qy     195 WLGADRFYLGYPAL-------GLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYIIDY 247
           |:      : | :|       | : : |  |  :  |:  ::  : : |   |  | |
Db     404 WV------IVYASLAKGIKTSGKVVYFTATFPYV-VLVILLIRGVTLPGAGAGIWYFITP 456

Qy     248 YGTRLTRLSITNETFRKTQLY 268
           :||  ::  :    ||::
```

```
Db    457 KWEKLTDATVWKDA--ATQIF 475


RESULT 13
US-08-700-013B-19
; Sequence 19, Application US/08700013B
; Patent No. 5919653
;  GENERAL INFORMATION:
;    APPLICANT:  Albert, Vivian R.
;    APPLICANT:  Kowalski, Leslie R.Z.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  McKelvy, Jeffrey F.
;    TITLE OF INVENTION:  Human Glycine Transporter
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/700,013B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  797 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    8-700-013B-19

  Query Match             5.5%;  Score 79;  DB 2;  Length 797;
  Best Local Similarity   23.9%;  Pred. No. 11;
  Matches   47;  Conservative   32;  Mismatches   80;  Indels   38;  Gaps   11;

Qy      87 FVSVTTGPWGAVATSAGGEESLKCED---LKVGQYICKD-PKINDATQEPVNCTNYTAHV 142
           ||||   |||:        |   |:|   | :   :  | |||           | |  :
Db     302 FVSVL--PWGSCNNPWNTPE---CKDKTKLLLDSCVISDHPKI--------QIKNSTFCM 348

Qy     143 SCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSY----KVAVALSLFLGWLGA 198
           : :|   :    |  |:|  :|:|  :||    :   |  |    :  :|| ||| |:
Db     349 TAYPNVTMVNFTSQANKTFVSGSE-EYFKYFVLKISAGIEYPGEIRWPLALCLFLAWV-- 406

Qy     199 DRFYLGYPAL-------GLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYIIDYYGTR 251
           : | :|       | : : |  |  :  |:  ::  : : |   |  | |     :
Db     406 ----IVYASLAKGIKTSGKVVYFTATFPYV-VLVILLIRGVTLPGAGAGIWYFITPKWEK 460

Qy     252 LTRLSITNETFRKTQLY 268
           ||  ::  :    ||::
Db     461 LTDATVWKDA--ATQIF 475


RESULT 14
US-08-414-657D-60
; Sequence 60, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-414-657D-60


  Query Match             5.5%;  Score 78.5;  DB 2;  Length 338;
  Best Local Similarity   27.2%;  Pred. No. 3.6;
  Matches   34;  Conservative   17;  Mismatches   47;  Indels   27;  Gaps    6;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |      ::   | :|| |    :  |   : ::      :  ::
Db     230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTEGQSSLTV----TMVTEEHY 285

Qy     157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
           || |    |::|        |:| | | :|| |  :|| |     ||      |    |
Db     286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----LAASLLC 333

Qy     210 LLKFC 214
           ||  |
Db     334 LLSKC 338


RESULT 15
US-08-356-340-2
; Sequence 2, Application US/08356340
; Patent No. 5608146
;  GENERAL INFORMATION:
;    APPLICANT:  FROMMER, Wolf-Bernd
;    APPLICANT:  RIESMEIER, Jorg
;    TITLE OF INVENTION:  DNA SEQUENCES WITH OLIGOSACCHARIDE
;    TITLE OF INVENTION:  TRANSPORTER, PLASMIDS, BACTERIA AND PLANTS CONTAINING A
;    TITLE OF INVENTION:  TRANSPORTER AS WELL AS A PROCESS FOR THE PREPARATION AND
```

```
;    TITLE OF INVENTION:  TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ostrolenk, Faber, Gerb & Soffen
;      STREET:  1180 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10036-8403
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,340
;      FILING DATE:  21-DEC-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/EP93/01604
;      FILING DATE:  22-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P42 20 759.2
;      FILING DATE:  24-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meilman, Edward A.
;      REGISTRATION NUMBER:  24,735
;      REFERENCE/DOCKET NUMBER:  P/951-106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 382-0700
;      TELEFAX:  (212) 382-0888
;      TELEX:  236925
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  525 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-356-340-2


  Query Match             5.3%;  Score 76;  DB 1;  Length 525;
  Best Local Similarity   28.9%;  Pred. No. 12;
  Matches   24;  Conservative    6;  Mismatches   17;  Indels   36;  Gaps    3;

Qy      40 LPFSLPLL------GGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTG 93
           : ||:||          ||| |  : |  :|   |                 |||||:|
Db     434 ITFSIPLALASIFSASSGSGQGLSLGVLNLAIVVPQ---------------MFVSVTSG 477

Qy      94 PW--------------GAVATSA 102
           ||              |||| :|
Db     478 PWDAMFGGGNLPAFVVGAVAATA 500


Search completed:  February 10, 2000, 17:01:46
Job time:  39 sec
```

OM of: US-09-060-609-2 to: EST:* out_format : pfs

Date: Feb 11, 2000 5:43 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09060609/runat_09022000_150438_25170/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09060609
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

[redacted]ch information block:
[redacted]ry: US-09-060-609-2
Query length: 269
Database: EST:*
Database sequences: 4538634
Database length: 1887831982
Search time (sec): 504.170000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est36:AI923178	+	1052.00	1998.61	1.5e-102	599	! AI923178 wn67b10.x1 NCI_CGAP_Lu
gb_est31:AI680904	+	1024.00	1945.46	1.3e-99	574	! AI680904 tx42f05.x1 NCI_CGAP_Lu
gb_est28:AI479764	+	988.00	1876.86	8.9e-96	556	! AI479764 tm69b04.x1 NCI_CGAP_Br
gb_est29:AI580361	+	957.00	1817.87	1.7e-92	537	! AI580361 tm47h02.x1 NCI_CGAP_Ki
gb_est31:AI674462	+	947.00	1798.86	2.0e-91	530	! AI674462 wc44e01.x1 NCI_CGAP_Pr
gb_est19:AA772225	+	924.50	1755.39	5.2e-89	548	! AA772225 ai41c01.s1 Soares_para
gb_est31:AI682204	+	905.00	1718.91	5.6e-87	507	! AI682204 wa71b06.x1 Soares_NFL_
gb_est26:AI349520	+	898.00	1705.70	3.0e-86	498	! AI349520 qp72g05.x1 Soares_feta
gb_est31:AI680969	+	872.00	1655.86	1.8e-83	500	! AI680969 tx35h07.x1 NCI_CGAP_Lu
gb_est6:N47594	+	828.50	1572.38	8.1e-79	507	! N47594 yy59a07.s1 Soares_multip
gb_est24:AI168073	+	805.50	1529.15	2.1e-76	471	! AI168073 oz89c02.x1 Soares_sene
gb_est25:AI299154	+	796.00	1511.74	1.9e-75	439	! AI299154 qm96f11.x1 NCI_CGAP_Lu
gb_est12:AA306979	+	791.00	1501.28	7.4e-75	475	! AA306979 EST178050 Colon carcin
gb_est22:AI038331	+	779.00	1478.95	1.3e-73	448	! AI038331 ox84a06.x1 Soares_sene
gb_est8:AA017855	-	777.00	1474.17	2.4e-73	488	! AA017855 mh47c09.r1 Soares mous
gb_est8:AA020016	-	777.00	1474.44	2.3e-73	476	! AA020016 mh48a12.r1 Soares mous
[redacted]est25:AI299329	+	767.00	1456.61	2.3e-72	423	! AI299329 qn22c06.x1 NCI_CGAP_Lu
[redacted]est25:AI284315	+	729.50	1382.05	3.2e-68	540	! AI284315 qj65c02.x1 NCI_CGAP_Ki
gb_est10:AA143062	+	718.50	1358.50	6.7e-67	674	! AA143062 zl49h02.s1 Soares_preg
gb_est24:AI175138	+	716.00	1357.01	8.1e-67	502	! AI175138 EST218660 Normalized r
gb_est25:AI245878	+	710.00	1348.20	2.5e-66	395	! AI245878 qk27f08.x1 NCI_CGAP_Ki
gb_est38:AW071521	+	708.00	1344.29	4.1e-66	398	! AW071521 xa36a07.x1 NCI_CGAP_Sa
gb_est17:AA639448	+	705.50	1336.76	1.1e-65	508	! AA639448 nq88g05.s1 NCI_CGAP_Co
gb_est19:AA757905	+	702.50	1333.26	1.7e-65	416	! AA757905 zg42h09.s1 Soares_pine
gb_est24:AI239974	+	702.50	1331.79	2.0e-65	474	! AI239974 qh44d02.x1 Soares_NFL_
gb_est11:AA218227	+	700.00	1323.66	5.8e-65	639	! AA218227 mv61g08.r1 Soares mous
gb_est31:AI693751	+	689.50	1307.20	4.8e-64	461	! AI693751 wd44h04.x1 Soares_NFL_
gb_est5:N34957	+	667.00	1264.04	1.2e-61	464	! N34957 yy50b11.s1 Soares_multip
gb_est21:AA943663	+	656.50	1244.09	1.6e-60	457	! AA943663 EST199162 Normalized r
gb_est6:D81362	+	655.00	1243.14	1.8e-60	385	! D81362 HUM159G08B Human fetal b
gb_est4:D61247	+	620.50	1176.16	9.5e-57	417	! D61247 HUM172D06B Clontech huma
gb_est4:R99199	+	613.50	1159.06	8.5e-56	580	! R99199 yq64d10.r1 Soares fetal
gb_est19:AA757694	+	611.00	1158.66	9.0e-56	392	! AA757694 zg38h03.s1 Soares_pine
gb_est13:AA345598	+	599.00	1135.96	1.7e-54	382	! AA345598 EST51648 Gall bladder
gb_est8:W84976	-	594.00	1127.18	5.1e-54	356	! W84976 mf42g02.r1 Soares mouse
gb_est18:AA711942	+	594.00	1125.33	6.5e-54	420	! AA711942 vu59h08.r1 Soares mous
gb_est24:AI176189	+	592.50	1122.51	9.3e-54	418	! AI176189 EST219770 Normalized r
gb_est36:AI903935	+	592.00	1124.47	7.2e-54	322	! AI903935 MR-BT038-080199-049 BT
gb_est36:AI903940	+	592.00	1124.44	7.3e-54	323	! AI903940 MR-BT038-080199-018 BT
gb_est36:AI903938	+	588.00	1116.77	1.9e-53	323	! AI903938 MR-BT038-080199-035 BT
gb_est24:AI194064	+	583.00	1106.79	7.0e-53	335	! AI194064 qe78d08.x1 Soares_feta
gb_est22:AI060271	+	547.00	1033.02	9.0e-49	515	! AI060271 UI-R-C1-11-f-12-0-UI.s
gb_est8:AA023218	-	541.00	1026.11	2.2e-48	342	! AA023218 mh69b08.r1 Soares m
gb_est36:AI887092	+	540.00	1025.66	2.3e-48	300	! AI887092 wl96f10.x1 NCI_CGAP
gb_est22:AI057115	-	533.00	1003.91	3.7e-47	632	! AI057115 oz23g10.x1 Soares_t

seq_name: gb_est36:AI923178

seq_documentation_block:
LOCUS AI923178 599 bp mRNA EST 02-SEP-1999
DEFINITION wn67b10.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2450491 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION AI923178
VERSION AI923178.1 GI:5659142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 599)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133491.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 457.
FEATURES Location/Qualifiers
source 1. .599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2450491"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went througb one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 148 a 127 c 146 g 176 t 2 others
ORIGIN

alignment_scores:
Quality: 1052.00 Length: 198
Ratio: 5.340 Gaps: 0
Percent Similarity: 99.495 Percent Identity: 98.990

alignment_block:
US-09-060-609-2 x AI923178 ..

Align seg 1/1 to: AI923178 from: 1 to: 599

64 AlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80

```
     1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAG 50

    80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    51 ACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

    97 laValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTC 150

   114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGl 130
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   151 AAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCA 200

   130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   201 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 250

   147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
       |||||:::||||||||||||||||||||||||||||||||||||||||||
   251 CACCCCACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

   164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGG 350

   180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   351 CTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 400

   197 lyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPhe 213
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   401 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTT 450

   214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIl 230
       |||||||||   ||||||||||||||||||||||||||||||||||||||
   451 TGCACTGTANGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTAT 500

   230 eSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIleAspT 247
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   501 TTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATT 550

   247 yrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThr 261
       |||||||||||||||||||||||||||||||||||||||||||
   551 ACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACA 594

seq_name: gb_est31:AI680904

documentation_block:
LOCUS       AI680904      574 bp    mRNA            EST       26-MAY-1999
DEFINITION  tx42f05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272257 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI680904
VERSION     AI680904.1  GI:4891086
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1134035.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2272257"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      148 a    118 c    136 g    172 t
ORIGIN


alignment_scores:
           Quality: 1024.00              Length:    191
             Ratio:  5.361                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-060-609-2 x AI680904    ..

 Align seg 1/1  to: AI680904  from: 1  to: 574

    71 ProSerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrpPh 87
       ||||||||||||||||||||||||||||||||||||||||||||||||||
     1 CCGTCTGCTCCGGAGGCCGTGACTGCCAGACTCGTTGGTGTCCTGTGGTT 50

    87 eValSerValThrThrGlyProTrpGlyAlaValAlaThrSerAlaGlyG 104
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    51 CGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGG 100

   104 lyGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCys 120
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   101 GCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGT 150

   121 LysAspProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAs 137
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   151 AAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAA 200

   137 nTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysA 154
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   201 CTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGG 250

   154 spSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePhe 170
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   251 ATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTC 300

   171 LysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVa 187
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   301 AAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGT 350

   187 lAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuG 204
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   351 CGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTG 400

   204 lyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGly 220
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   401 GATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGA 450
```

```
     221 IleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleValGlyPr 237
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 ATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACC 500

     237 oSerAspGlySerSerTyrIleIleAspTyrTyrGlyThrArgLeuThrA 254
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 TTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAA 550

     254 rgLeuSerIleThrAsnGluThr 261
         |||||||||||||||||||||
     551 GACTGAGTATTACTAATGAAACA 573

seq_name: gb_est28:AI479764

seq_documentation_block:
LOCUS       AI479764      556 bp    mRNA            EST       14-APR-1999
DEFINITION  tm69b04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163343 3',
            mRNA sequence.
ACCESSION   AI479764
VERSION     AI479764.1  GI:4372932
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188819.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 614   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 456.
FEATURES            Location/Qualifiers
     source         1. .556
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2163343"
                    /clone_lib="NCI_CGAP_Brn25"
                    /tissue_type="anaplastic oligodendroglioma"
                    /lab_host="DH10B"
                    /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTT
                    T 3']; double-stranded cDNA was ligated to Eco RI
                    adaptors (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of the modified pT7T3 vector.
                    Library is normalized, and was constructed by Bento
                    Soares and M.Fatima Bonaldo."
BASE COUNT      148 a    109 c    130 g    168 t      1 others
ORIGIN


alignment_scores:
            Quality:  988.00            Length:    185
              Ratio:  5.370               Gaps:      0
 Percent Similarity:  99.459   Percent Identity: 99.459
```

```
alignment_block:
 US-09-060-609-2 x AI479764    ..

 Align seg 1/1  to: AI479764  from: 1  to: 556

      79 AlaArgLeuValGlyValLeuTrpPheValSerValThrThrGlyProTr 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 GCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTG 51

      95 pGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluA 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 GGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGG 101

     112 spLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAla 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 ACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCT 151

     129 ThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 ACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTT 201

     145 eProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisP 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATT 251

     162 heThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVal 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTA 301

     179 AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTr 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 AATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATG 351

     195 pLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuL 212
         |||||||||||||||   ||||||||||||||||||||||||||||||||
     352 GTTGGGAGCAGATCGATTNTACCTTGGATACCCTGCTTTGGGTTTGTTAA 401

     212 ysPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIle 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 AGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATT 451

     229 LeuIleSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIl 245
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 CTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTAT 501

     245 eAspTyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrP 262
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     502 AGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACAT 551

     262 heArg 263
         |||||
     552 TTAGA 556

seq_name: gb_est29:AI580361

seq_documentation_block:
LOCUS       AI580361      537 bp    mRNA            EST       05-APR-1999
DEFINITION  tm47h02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2161299 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI580361
VERSION     AI580361.1  GI:4564737
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949103.
```

```
          Contact: Robert Strausberg, Ph.D.
          Tel: (301) 496-1550
          Email: Robert_Strausberg@nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html

          Seq primer: -40UP from Gibco
          High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2161299"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      124 a    117 c    139 g    156 t      1 others
ORIGIN


alignment_scores:
           Quality:  957.00            Length:    179
             Ratio:  5.376               Gaps:      0
 Percent Similarity:  99.441   Percent Identity: 98.883

alignment_block:
 US-09-060-609-2 x AI580361   ..

 Align seg 1/1  to: AI580361  from: 1  to: 537

      64 AlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAG 50

      80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

      97 laValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTC 150

     114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGl 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 AAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCA 200

     130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 250

     147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

     164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGG 350

     180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 400

     197 lyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPhe 213
         ||||||||||||||||||||||||||||||||  ||||||||||||||||
     401 GAGCAGATCGATTTTACCTTGGATACCCTGCTNTGGGTTTGTTAAAGTTT 450

     214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIl 230
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 TGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTAT 500

     230 eSerMetGlnIleValGlyProSerAspGlySerSer 242
         ||||||||||||||||||||||||||||||:::|||
     501 TTCAATGCAGATTGTTGGACCTTCAGATGGGAAGAGT 537

seq_name: gb_est31:AI674462

seq_documentation_block:
LOCUS       AI674462      530 bp    mRNA            EST       19-MAY-1999
DEFINITION  wc44e01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321496 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI674462
VERSION     AI674462.1  GI:4874942
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948555.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2321496"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      123 a    116 c    133 g    157 t      1 others
ORIGIN


alignment_scores:
           Quality:  947.00            Length:    176
```

```
            Ratio:  5.411              Gaps:      0
 Percent Similarity:  99.432   Percent Identity: 99.432

alignment_block:
 US-09-060-609-2 x AI674462   ..

 Align seg 1/1  to: AI674462  from: 1  to: 530

      66 AlaTrpProSerGlyProSerAlaProGluAlaValThrAlaArgLeuVa 82
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTTACGGCCAGACTCGT 52

      82 lGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyAlaValA 99
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 TGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTG 102

      99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLysVal 115
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTG 152

     116 GlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluPr 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 GGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACC 202

     132 oValAsnCysThrAsnTyrThrAlaHisValSerCysPheProAlaProA 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 AGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCA 252

     149 snIleThrCysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsn 165
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 ACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAAC 302

     166 GluValGlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrSe 182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 GAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTC 352

     182 rTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaA 199
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAG 402

     199 spArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThr 215
         ||||||||||||||||||||||||||||||||||||||||   ||||||
     403 ATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGNTTTGCACT 452

     216 ValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMe 232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 GTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAAT 502

     232 tGlnIleValGlyProSerAspGlySer 241
         |||||||||||||||||||||||||||
     503 GCAGATTGTTGGACCTTCAGATGGAAGT 530

seq_name: gb_est19:AA772225

seq_documentation_block:
LOCUS       AA772225      548 bp    mRNA            EST       31-DEC-1998
DEFINITION  ai41c01.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            1359552 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;,
            mRNA sequence.
ACCESSION   AA772225
VERSION     AA772225.1  GI:2824008
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811076.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Insert Length: 978   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="1359552"
                     /clone_lib="Soares_parathyroid_tumor_NbHPA"
                     /tissue_type="parathyroid tumor"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: parathyroid gland; Vector: pT7T3D
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer
                     [5'-
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTT
                     T-3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                     adenomas was kindly provided by Dr. Stephen Marx, National
                     Institute of Diabetes and Digestive and Kidney Diseases,
                     NIH."
BASE COUNT      146 a    106 c    126 g    169 t      1 others
ORIGIN


alignment_scores:
            Quality:  924.50             Length:    184
              Ratio:  5.136                Gaps:      3
 Percent Similarity:  97.826   Percent Identity: 96.196

alignment_block:
 US-09-060-609-2 x AA772225   ..

 Align seg 1/1  to: AA772225  from: 1  to: 548

      81 LeuValGlyValLeuTrpPheVal.SerValThrThrGlyProTrpGlyA 97
         |||||||||||||||||||||||| :::||||||||||||||||||||||
       2 CTCGTTGGTGTCCTGTGGTTCGTCTGCAGTCACTACAGGACCCTGGGGG. 50

      97 laVal.AlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLe 113
           ::: :::|||||||||||||||   ||||||||||||||||||||||
      51 ..CTGTTGCACCTCCGCCGGGGGCANGGAGTCGCTTAAGTGCGAGGACCT 98

     113 uLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      99 CAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGC 148

     130 lnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPhePro 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     149 AAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCA 198

     147 AlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheTh 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     199 GCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTAC 248
```

```
    163 rGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    249 TGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATG 298

    180 lyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeu 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    299 GCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTG 348

    197 GlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPh 213
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    349 GGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTT 398

    213 eCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuI 230
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    399 TTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTA 448

    230 leSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIleAsp 246
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    449 TTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGAT 498

    247 TyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrPhe 262
        ||||||||||||||||||||||||||||||||||||||||||||||||
    499 TACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACATTT 546

seq_name: gb_est31:AI682204

seq_documentation_block:
LOCUS       AI682204      507 bp    mRNA            EST       26-MAY-1999
DEFINITION  wa71b06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2301587 3', mRNA sequence.
ACCESSION   AI682204
VERSION     AI682204.1  GI:4892386
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1136077.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2301587"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      120 a    109 c    125 g    152 t      1 others
ORIGIN

alignment_scores:
           Quality:  905.00            Length:    169
             Ratio:  5.387               Gaps:      0
 Percent Similarity:  99.408   Percent Identity: 99.408

alignment_block:
 US-09-060-609-2 x AI682204   ..

 Align seg 1/1  to: AI682204  from: 1  to: 507

      70 GlyProSerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTr 86
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTG 50

      86 pPheValSerValThrThrGlyProTrpGlyAlaValAlaThrSerAlaG 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCG 100

     103 lyGlyGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIle 119
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATT 150

     120 CysLysAspProLysIleAsnAspAlaThrGlnGluProValAsnCysTh 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 TGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTAC 200

     136 rAsnTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysL 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTA 250

     153 ysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhe 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

     170 PheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAl 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGC 350

     186 aValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrL 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 AGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACC 400

     203 euGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCys 219
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 TTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTANGGTTTTGT 450

     220 GlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleValGl 236
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 GGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGG 500

     236 yProSer 238
         |||||||
     501 ACCTTCA 507

seq_name: gb_est26:AI349520

seq_documentation_block:
LOCUS       AI349520      498 bp    mRNA            EST       03-FEB-1999
DEFINITION  qp72g05.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1928600 3', mRNA sequence.
ACCESSION   AI349520
VERSION     AI349520.1  GI:4086726
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
          Tel: (301) 496-1550
          Email: Robert_Strausberg@nih.gov
          This clone is available royalty-free through LLNL ; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert Length: 1009   Std Error: 0.00
          Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1928600"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT      113 a    112 c    128 g    144 t      1 others
ORIGIN


alignment_scores:
           Quality:  898.00            Length:    166
             Ratio:  5.442               Gaps:      0
 Percent Similarity:  99.398   Percent Identity: 99.398

alignment_block:
 US-09-060-609-2 x AI349520   ..

 Align seg 1/1  to: AI349520  from: 1  to: 498

     64 AlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAG 50

     80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

     97 laValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTC 150

    114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGl 130
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 AAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCA 200

    130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 250

    147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

    164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGG 350

    180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 CTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 400

    197 lyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPhe 213
        |||||||||||||||||||||||||||||||   ||||||||||||||||
    401 GAGCAGATCGATTTTACCTTGGATACCCTGCTNTGGGTTTGTTAAAGTTT 450

    214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeu 229
        |||||||||||||||||||||||||||||||||||||||||||||||||
    451 TGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 498

seq_name: gb_est31:AI680969

seq_documentation_block:
LOCUS       AI680969      500 bp    mRNA            EST       26-MAY-1999
DEFINITION  tx35h07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2271613 3',
            mRNA sequence.
ACCESSION   AI680969
VERSION     AI680969.1  GI:4891151
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135321.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2271613"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      114 a    112 c    125 g    145 t      4 others
ORIGIN


alignment_scores:
           Quality:  872.00            Length:    166
             Ratio:  5.383               Gaps:      0
 Percent Similarity:  97.590   Percent Identity: 96.988

alignment_block:
 US-09-060-609-2 x AI680969   ..

 Align seg 1/1  to: AI680969  from: 1  to: 500

     64 AlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80
        ||||||||||||||||||||||||||||||||||||:::|||||||||
```

```
      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCTTGACGGCCAG 50

     80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

     97 laValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTC 150

    114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGl 130
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 AAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCA 200

    130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 250

    147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

    164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGG 350

    180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 CTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGN 400

    197 lyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPhe 213
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTT 450

    214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeu 229
        |||||||||      ||||||||||||||||||||   |||||||||
    451 TGCACTGTANNGGTTTGTGGAATTGGGAGCCTAATTGANTTCATTCTT 498

seq_name: gb_est6:N47594

seq_documentation_block:
LOCUS       N47594        507 bp    mRNA            EST       14-FEB-1996
DEFINITION  yy59a07.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA
            clone IMAGE:277812 3', mRNA sequence.
ACCESSION   N47594
VERSION     N47594.1  GI:1188760
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:802394.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyT not found
            Seq primer: m13 -40 forward
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /db_xref="GDB:3896188"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:277812"
                     /clone_lib="Soares_multiple_sclerosis_2NbHMSP"
                     /sex="male"
                     /tissue_type="multiple sclerosis lesions"
                     /dev_stage="Age 46"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                     lesions from one patient was kindly provided by Dr. Kevin
                     G. Becker (NINDS/NIH). "
BASE COUNT      117 a    109 c    124 g    148 t      9 others
ORIGIN

alignment_scores:
            Quality:  828.50            Length:    171
              Ratio:  5.114               Gaps:      3
 Percent Similarity:  94.737   Percent Identity: 93.567

alignment_block:
 US-09-060-609-2 x N47594   ..

 Align seg 1/1  to: N47594  from: 1  to: 507

     64 AlaAlaAlaTrpProSerGlyProSerAlaProGlu.AlaValThrAlaA 80
        |||:::||||||   :::||||||||||||||||||          |||
      1 GCGNCCGCCTGGNACNTNGGTCCGTCTGCTCCGGAGCCGT...GACGCAG 47

     80 rgLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGly 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     48 NACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGG 97

     97 AlaValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLe 113
        ||||||||||||||||||||      ||||||||||||||||||||||||
     98 GCTGTTGCCACCTCCGCCGGGGNNAGNGAGTCGCTTAAGTGCGAGGACCT 147

    113 uLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    148 CAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGC 197

    130 lnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPhePro 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    198 AAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCA 247

    147 AlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheTh 163
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    248 GCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTAC 297

    163 rGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    298 TGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATG 347

    180 lyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeu 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    348 GCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTG 397

    197 GlyAlaAspArgPheTyrLeuGlyTyrProAlaLeu.GlyLeuLeuLysP 213
        |||||||||||||||||||||||||||||||||||| ||||||||||||
    398 GGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGGTTTGTTAAAGT 447

    213 heCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeu 229
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 448 TTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 497

 230 IleSerMet 232
     |||||||||
 498 ATTTCAATG 506

seq_name: gb_est24:AI168073

seq_documentation_block:
LOCUS       AI168073      471 bp    mRNA            EST       05-OCT-1998
DEFINITION  oz89c02.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
            clone IMAGE:1682498 3', mRNA sequence.
ACCESSION   AI168073
VERSION     AI168073.1  GI:3701243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2282128.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1682498"
                     /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      111 a    110 c    120 g    130 t
ORIGIN


alignment_scores:
            Quality:  805.50            Length:    182
              Ratio:  5.197               Gaps:      2
 Percent Similarity:  85.165   Percent Identity: 85.165

alignment_block:
 US-09-060-609-2 x AI168073   ..

 Align seg 1/1  to: AI168073  from: 1  to: 471

      64 AlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAG 50

      80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97
         |||||||||||||||||||||||||||||||||||||||||||||||||
      51 ACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

      97 laValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTC 150

     114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGl 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 AAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCA 200

     130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 250

     147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
         |||||||||||||||||||||||||||||||||||||||||||||||||
     251 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

     164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGG 350

     180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 400

     197 lyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPhe 213
         |||||||||||||||||||||||||||||||
     401 GAGCAGATCGATTTTACCTTGGATACCCTGCT..................... 432

     214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIl 230

     432 ......................................................... 432

     230 eSerMetGln.IleValGlyProSerAspGlySerSerTyrIle 244
                   ||||||||||||||||||||||||||||||||||||
     433 .......TTGGATTGTTGGACCTTCAGATGGAAGTAGTTACATT 469

seq_name: gb_est25:AI299154

seq_documentation_block:
LOCUS       AI299154      439 bp    mRNA            EST       29-JAN-1999
DEFINITION  qm96f11.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1896621 3',
            mRNA sequence.
ACCESSION   AI299154
VERSION     AI299154.1  GI:3958808
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 966   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone="IMAGE:1896621"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      105 a    100 c    110 g    124 t
ORIGIN


alignment_scores:
           Quality:  796.00            Length:    146
             Ratio:  5.452               Gaps:      0
rcent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-060-609-2 x AI299154   ..

 Align seg 1/1  to: AI299154  from: 1  to: 439

      70 GlyProSerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTr 86
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTG 50

      86 pPheValSerValThrThrGlyProTrpGlyAlaValAlaThrSerAlaG 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCG 100

     103 lyGlyGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIle 119
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATT 150

     120 CysLysAspProLysIleAsnAspAlaThrGlnGluProValAsnCysTh 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 TGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTAC 200

     136 rAsnTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysL 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTA 250

     153 ysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhe 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

     170 PheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAl 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGC 350

     186 aValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrL 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 AGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACC 400

     203 euGlyTyrProAlaLeuGlyLeuLeuLysPheCysThr 215
         |||||||||||||||||||||||||||||||||||||
     401 TTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACT 438

seq_name: gb_est12:AA306979

seq_documentation_block:
LOCUS       AA306979      475 bp    mRNA            EST       18-APR-1997
DEFINITION  EST178050 Colon carcinoma (HCC) cell line Homo sapiens cDNA 5' end,
            mRNA sequence.
ACCESSION   AA306979
VERSION     AA306979.1  GI:1959469
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., ,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1288147.
            Other_ESTs: THC195202
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):108677"
                     /db_xref="taxon:9606"
                     /clone_lib="Colon carcinoma (HCC) cell line"
                     /tissue_type="colon"
                     /cell_type="KM12C"
                     /cell_line="KM12C(HCC)-parental human colon
                     carcinoma;Dukes B2"
                     /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      108 a    107 c    119 g    135 t      6 others
ORIGIN


alignment_scores:
           Quality:  791.00            Length:    159
             Ratio:  5.136               Gaps:      1
 Percent Similarity:  96.855   Percent Identity: 96.226

alignment_block:
 US-09-060-609-2 x AA306979   ..

 Align seg 1/1  to: AA306979  from: 1  to: 475

      64 AlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80
         ||||||||||||||:::|||||||||||||||||||||||||||   ||
       2 GCGGCCGCCTGGCCTNCTGGTCCGTCTGCTCCGGAGGCCGTGACGGC.AG 50

      80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGG 100

      97 laValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113
         |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     101 CTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTC 150

     114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGl 130
         ||||||||||||||||||   ||||||||||||||||||   ||||||||
     151 AAAGTGGGACAATATATTTNTAAAGATCCAAAAATAAATNACGCTACGCA 200

     130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAG 250

     147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

     164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180
         |||||||||      ||||||||||||||||||||||||||||||||||
     301 GGGAACGAANTNGNTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGG 350

     180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGG 400

     197 lyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLys.Ph 213
         |||||||||||||||||||||||||||||||||||||||||||||||| ||
     401 GAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTT 450

     213 eCysThrValGlyPheCysGlyIle 221
         |||||||||||||||||||||||||
     451 TTGCACTGTAGGGTTTTGTGGAATT 475

seq_name: gb_est22:AI038331

seq_documentation_block:
LOCUS       AI038331      448 bp    mRNA            EST       24-SEP-1998
DEFINITION  ox84a06.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
            clone IMAGE:1662994 3', mRNA sequence.
ACCESSION   AI038331
VERSION     AI038331.1  GI:3277525
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
  COMMENT   On May 9, 1995 this sequence version replaced gi:802984.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1099   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1662994"
                     /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      118 a     92 c    105 g    133 t
ORIGIN


alignment_scores:
            Quality:  779.00            Length:    146
              Ratio:  5.372               Gaps:      0
 Percent Similarity:  99.315   Percent Identity: 98.630

alignment_block:
 US-09-060-609-2 x AI038331   ..

 Align seg 1/1  to: AI038331  from: 1  to: 448

      88 ValSerValThrThrGlyProTrpGlyAlaValAlaThrSerAlaGlyGl 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       5 GTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGG 54

     104 yGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCysL 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      55 CGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTA 104

     121 ysAspProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsn 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     105 AAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAAC 154

     138 TyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAs 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     155 TACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGA 204

     154 pSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheL 171
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 TTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCA 254

     171 ysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVal 187
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 AGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTC 304

     188 AlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGl 204
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     305 GCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGG 354

     204 yTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyI 221
         ||||||||||||||||||||||||||||||||||   :::|||||||||
     355 ATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTATGGTTATGTGGAA 404

     221 leGlySerLeuIleAspPheIleLeuIleSerMetGln 233
         |||||||||||||||||||||||||||||||||||||
     405 TTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAG 442

seq_name: gb_est8:AA017855

seq_documentation_block:
LOCUS       AA017855      488 bp    mRNA            EST       21-JAN-1997
DEFINITION  mh47c09.r1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA
            clone IMAGE:445648 5' similar to WP:C02F5.3 CE00039 GTP-BINDING
            PROTEIN ;, mRNA sequence.
ACCESSION   AA017855
VERSION     AA017855.1  GI:1481174
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
```

```
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785679.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:270984
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:445648"
                     /clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
                     /sex="unknown"
                     /tissue_type="placenta"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      150 a    101 c     87 g    150 t
ORIGIN


alignment_scores:
            Quality:  777.00              Length:    154
              Ratio:  5.112                 Gaps:      0
 Percent Similarity:  98.701   Percent Identity: 95.455

alignment_block:
 09-060-609-2 x AA017855/rev   ..

 Align seg 1/1  to reverse of: AA017855  from: 1  to: 488

    116 GlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluPr 132
        ||||||||||||||||||:::|||||||||||||||||   ||||||||||
    471 GGACAATATATTTGTAAAGAACCAAAAATAAATGATCG.ACGCAAGAACC 423

    132 oValAsnCysThrAsnTyrThrAlaHisValSerCysPheProAlaProA 149
        |||||||||||||||||||||||||||||:::|||||||||||||||:
    422 AGTTAATTGTACAAACTACACAGCTCATGTTCAATGTTTTCCAGCACCCA 373

    149 snIleThrCysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsn 165
        ::|||||||||||||||   |||||||||||||||||||||||||:::
    372 AAATAACTTGTAAGGATTTGAGTGGTAATGAAACACATTTTACTGGAAGT 323

    166 GluValGlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrSe 182
        ||||||||||||:::|||||||||||||||||||||||||||||||||||
    322 GAAGTCGGTTTTCTCAAGCCCATATCTTGCCGAAATGTGAATGGCTATTC 273

    182 rTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaA 199
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    272 GTACAAAGTGGCAGTTGCATTATCTCTCTTTTTGGGATGGCTGGGAGCAG 223

    199 spArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThr 215
        |||||||||||||||||||||||||||||||||||||||||||||||||||

    222 ATCGATTTTACCTCGGATATCCTGCCTTAGGCTTGTTAAAATTTTGCACC 173

    216 ValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMe 232
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    172 GTAGGATTTTGCGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAAT 123

    232 tGlnIleValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrG 249
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    122 GCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGACTATTATG 73

    249 lyThrArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThr 265
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     72 GAACCAGGCTTACAAGACTCAGCATTACTAATGAAACATTTAGAAAAACC 23

    266 GlnLeuTyrPro 269
        ||||||||||||
     22 CAGCTGTACCCA 11
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 04:15:42 ; Search time 13.34 Seconds
(without alignments)
477.628 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 1439
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

[redacted]base : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1439	100.0	269	1	W94291	Human beta-amyloid
2	748	52.0	139	1	Y12358	Human 5' EST secre
3	667.5	46.4	148	1	Y12426	Human 5' EST secre
4	88	6.1	764	1	W94360	Human Factor B ana
5	86	6.0	764	1	W94365	Human Factor B ana
6	86	6.0	764	1	W94372	Human Factor B ana
7	85	5.9	338	1	W05153	Rat LAMP residues
8	85	5.9	338	1	W05154	Rat LAMP residues
9	85	5.9	310	1	W05156	Rat mature LAMP. D
10	85	5.9	764	1	W94361	Human Factor B ana
11	84	5.8	764	1	W94366	Human Factor B ana
12	84	5.8	764	1	W94373	Human Factor B ana
13	83.5	5.8	763	1	W94362	Human Factor B ana
14	83	5.8	764	1	W94383	Human Factor B ana
15	83	5.8	764	1	W94384	Human Factor B ana
16	83	5.8	764	1	W94385	Human Factor B ana
17	83	5.8	764	1	W94386	Human Factor B ana
18	83	5.8	764	1	W94387	Human Factor B ana
19	83	5.8	764	1	W94388	Human Factor B ana
20	83	5.8	764	1	W94389	Human Factor B ana
21	83	5.8	762	1	W94390	Human Factor B ana
22	83	5.8	764	1	W94352	Human Factor B. Fa
23	83	5.8	764	1	W94353	Human Factor B ana
24	83	5.8	764	1	W94354	Human Factor B ana
25	83	5.8	764	1	W94355	Human Factor B ana
26	83	5.8	764	1	W94356	Human Factor B ana
27	83	5.8	764	1	W94359	Human Factor B ana
28	83	5.8	764	1	W94377	Human Factor B ana
29	83	5.8	764	1	W94378	Human Factor B ana
30	83	5.8	764	1	W94379	Human Factor B ana
31	83	5.8	764	1	W94380	Human Factor B ana
32	83	5.8	764	1	W94381	Human Factor B ana
33	83	5.8	763	1	W94382	Human Factor B ana
34	82	5.7	409	1	W43449	Tobacco laccase cl
35	82	5.7	764	1	W94364	Human Factor B ana
36	82	5.7	764	1	W94376	Human Factor B ana
37	81	5.6	797	1	W56378	Human GlyT-2 trans
38	80.5	5.6	325	1	W05152	Human LAMP residue
39	80.5	5.6	304	1	W05155	Human mature LAMP.
40	80.5	5.6	1237	1	W44729	Chicken protein ty
41	80.5	5.6	456	1	W69555	Human equilibrativ
42	80.5	5.6	373	1	W64550	Human fibrosarcoma
43	80.5	5.6	1237	1	W89347	Chicken transmembr
44	80.5	5.6	763	1	W94367	Human Factor B ana
45	80.5	5.6	763	1	W94358	Human Factor B ana

ALIGNMENTS

RESULT 1
W94291
ID W94291 standard; Protein; 269 AA.
AC W94291;
DT 27-APR-1999 (first entry)
DE Human beta-amyloid peptide-binding protein (BBP).
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;
KW human; Alzheimer's disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 68. .269
FT /note= "specifically claimed fragment having
FT beta-amyloid peptide binding activity"
PN WO9846636-A2.
PD 22-OCT-1998.
PF 14-APR-1998; U07462.
PR 16-APR-1997; US-064583.
PA (AMHP) AMERICAN HOME PROD CORP.
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,
PI Walker SG;
DR WPI; 99-080736/07.
DR N-PSDB; X05735.
PT Polynucleotide encoding beta-amyloid peptide binding protein - used
PT to identify inhibitors of beta-amyloid peptide for treating
PT Alzheimer's disease
PS Claim 7; Pages 43-44; 59pp; English.
CC The present sequence represents a beta-amyloid peptide binding protein
CC (BBP). The polynucleotide comprising the entire BBP nucleotide sequence
CC of clone BBP1-fl is deposited under the accession number ATCC 98617. The
CC polynucleotide comprising a fragment of BBP (nucleotides 202-807 of the
CC full length BBP) of clone pEK196 is deposited as ATCC 98399. Host cells
CC transformed with a vector comprising the BBP nucleic acid are used for
CC the recombinant production of the protein. The protein can be used in a
CC method for diagnosing a disease characterised by aberrant expression of
CC human beta-amyloid protein (BAP). The protein can also be used in a
CC method for screening for compounds which regulate expression of a BAP
CC binding protein. The proteins, antibodies and identified compounds can be
CC used in the treatment or prevention of Alzheimer's disease.
SQ Sequence 269 AA;

Query Match 100.0%; Score 1439; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 5.1e-137;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60

Qy      61 SKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC 120

Qy     121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180

Qy     181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240

Qy     241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269
```

```
          ||||||||||||||||||||||||||||
Db    241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269


RESULT   2
Y12358
ID   Y12358 standard; Protein; 139 AA.
AC   Y12358;
DT   17-JUN-1999  (first entry)
DE   Human 5' EST secreted protein SEQ ID NO:389.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
[redacted]
PF   31-JUL-1998; IB1222.
[redacted]
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   N-PSDB; X41191.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 27; Page 714-715; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   139 AA;

[redacted]

  Query Match             52.0%;  Score 748;  DB 1;  Length 139;
  Best Local Similarity   99.3%;  Pred. No. 6.5e-68;
  Matches  138;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      63 MAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 122
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAAAWXSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 60

Qy     123 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 120

Qy     183 YKVAVALSLFLGWLGADRF 201
           |||||||||||||||||||
Db     121 YKVAVALSLFLGWLGADRF 139


RESULT   3
Y12426
ID   Y12426 standard; Protein; 148 AA.
AC   Y12426;
DT   17-JUN-1999  (first entry)
DE   Human 5' EST secreted protein SEQ ID NO:457.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   N-PSDB; X41259.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 27; Page 763-764; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   148 AA;


  Query Match             46.4%;  Score 667.5;  DB 1;  Length 148;
  Best Local Similarity   83.3%;  Pred. No. 9e-60;
  Matches  125;  Conservative    4;  Mismatches   18;  Indels    3;  Gaps    2;

Qy      63 MAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 122
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAAAWPSGPXAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 60

Qy     123 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 182
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      61 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDXSGNETHFTGNEVGFFKPISCRNVNGYS 120

Qy     183 YKVAVALSLFLGWLGADRFYLGY-PALGLL 211
           |     :|    |:   |  | : |  | :
Db     121 YXXQXXVS--FSWMVGSRSILPWIPCFGFV 148


RESULT   4
W94360
ID   W94360 standard; Protein; 764 AA.
AC   W94360;
DT   13-APR-1999  (first entry)
DE   Human Factor B analogue #7.
KW   Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW   stroke; transplant rejection suppression; myocardial infarction.
OS   Homo sapiens.
OS   Synthetic.
PN   US5869615-A.
PD   09-FEB-1999.
PF   03-JAN-1994; 177109.
PR   03-JAN-1994; US-177109.
PA   (UNIW ) UNIV WASHINGTON.
PI   Hourcade DE, Oglesby TJ;
DR   WPI; 99-152877/13.
PT   Factor B analogues - with modified complement-mediated activity
```

PS Claim 18; Column -; 53pp; English.
CC The present sequence represents a human Factor B analogue. The present
CC invention describes human Factor B analogues (see W94353 to W94390) that
CC exhibit modified complement-mediated activity in vitro and consist of
CC the sequence given in W94352, of 764 amino acids, with one or more amino
CC acids in a short consensus repeat (SCR) domain or a von Willebrand
CC Factor (vWF) domain replaced by one or more amino acids from a
CC corresponding region of an SCR or vWF domain of a second protein, or
CC in which a protease domain of native Factor B is replaced by a protease
CC domain of C2. Complement protein analogues having altered activity can
CC be used to control activation of the complement system and are so are
CC useful for treating disorders associated with abnormal and/or excessive
CC complement activation e.g. for treating autoimmune diseases, suppressing
CC transplant rejection or reducing tissue damage associated with myocardial
CC infarction and stroke.
CC N.B. The present sequence is not given in the specification but is
CC derived from the sequence in W94352 as specified in the claims.
SQ Sequence 764 AA;

Query Match 6.1%; Score 88; DB 1; Length 764;
Best Local Similarity 24.6%; Pred. No. 1.9;
Matches 50; Conservative 22; Mismatches 69; Indels 62; Gaps 12;

```
Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG------------GSGSGEKVSV----------- 61
           |: | |       | |:|| | || ||            ||  |  |  :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
           | :  : |||  |   :  | |          :|| |  : |           :: |
Db      56 SLLTYSCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183
```

RESULT 5
W94365
ID W94365 standard; Protein; 764 AA.
AC W94365;
DT 13-APR-1999 (first entry)
DE Human Factor B analogue #12.
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW stroke; transplant rejection suppression; myocardial infarction.
OS Homo sapiens.
OS Synthetic.
PN US5869615-A.
PD 09-FEB-1999.
PF 03-JAN-1994; 177109.
PR 03-JAN-1994; US-177109.
PA (UNIW) UNIV WASHINGTON.
PI Hourcade DE, Oglesby TJ;
DR WPI; 99-152877/13.
PT Factor B analogues - with modified complement-mediated activity
PS Claim 18; Column -; 53pp; English.
CC The present sequence represents a human Factor B analogue. The present
CC invention describes human Factor B analogues (see W94353 to W94390) that
CC exhibit modified complement-mediated activity in vitro and consist of
CC the sequence given in W94352, of 764 amino acids, with one or more amino
CC acids in a short consensus repeat (SCR) domain or a von Willebrand
CC Factor (vWF) domain replaced by one or more amino acids from a
CC corresponding region of an SCR or vWF domain of a second protein, or
CC in which a protease domain of native Factor B is replaced by a protease
CC domain of C2. Complement protein analogues having altered activity can
CC be used to control activation of the complement system and are so are
CC useful for treating disorders associated with abnormal and/or excessive
CC complement activation e.g. for treating autoimmune diseases, suppressing
CC transplant rejection or reducing tissue damage associated with myocardial
CC infarction and stroke.
CC N.B. The present sequence is not given in the specification but is
CC derived from the sequence in W94352 as specified in the claims.
SQ Sequence 764 AA;

Query Match 6.0%; Score 86; DB 1; Length 764;
Best Local Similarity 24.6%; Pred. No. 3;
Matches 50; Conservative 21; Mismatches 70; Indels 62; Gaps 12;

```
Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG------------GSGSGEKVSV----------- 61
           |: | |       | |:|| | || ||            ||  |  |  :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVAT------SAGGEESLKC-- 111
             :   |||  |   :  | |          :|| |  : |       |     :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQRSLSKAVCRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183
```

RESULT 6
W94372
ID W94372 standard; Protein; 764 AA.
AC W94372;
DT 13-APR-1999 (first entry)
DE Human Factor B analogue #19.
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW stroke; transplant rejection suppression; myocardial infarction.
OS Homo sapiens.
OS Synthetic.
PN US5869615-A.
PD 09-FEB-1999.
PF 03-JAN-1994; 177109.
PR 03-JAN-1994; US-177109.
PA (UNIW) UNIV WASHINGTON.
PI Hourcade DE, Oglesby TJ;
DR WPI; 99-152877/13.
PT Factor B analogues - with modified complement-mediated activity
PS Claim 18; Column -; 53pp; English.
CC The present sequence represents a human Factor B analogue. The present
CC invention describes human Factor B analogues (see W94353 to W94390) that
CC exhibit modified complement-mediated activity in vitro and consist of
CC the sequence given in W94352, of 764 amino acids, with one or more amino
CC acids in a short consensus repeat (SCR) domain or a von Willebrand
CC Factor (vWF) domain replaced by one or more amino acids from a
CC corresponding region of an SCR or vWF domain of a second protein, or
CC in which a protease domain of native Factor B is replaced by a protease
CC domain of C2. Complement protein analogues having altered activity can
CC be used to control activation of the complement system and are so are
CC useful for treating disorders associated with abnormal and/or excessive
CC complement activation e.g. for treating autoimmune diseases, suppressing
CC transplant rejection or reducing tissue damage associated with myocardial
CC infarction and stroke.
CC N.B. The present sequence is not given in the specification but is
CC derived from the sequence in W94352 as specified in the claims.
SQ Sequence 764 AA;

Query Match 6.0%; Score 86; DB 1; Length 764;
Best Local Similarity 24.1%; Pred. No. 3;
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

```
Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG------------GSGSGEKVSV----------- 61
           |: | |       | |:|| | || ||            ||  |  |  :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55
```

```
Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
           :    |||  |   :  | |          :|| |  : |           :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCRPNGMWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


RESULT  7
W05153
ID   W05153 standard; Protein; 338 AA.
AC   W05153;
DT   22-MAY-1997  (first entry)
DE   Rat LAMP residues 1-332.
KW   limbic system associated membrane protein; LAMP; self binding domain;
KW   antibody-like; cell surface adhesion protein; neuron;
KW   monoclonal antibody; 2G9; growth; differentiation; epilepsy;
KW   Alzheimer's disease; schizophrenia; neural stem cell.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   peptide         1. .28
FT                   /note= "Signal peptide"
FT   protein         29. .338
FT                   /note= "Mature protein"
PN   WO9630052-A1.
PD   03-OCT-1996.
PF   29-MAR-1996; U04397.
PR   31-MAR-1995; US-414657.
PA   (UMDJ-) UMDNJ UNIV NEW JERSEYS HEALTH SCI.
PI   Fischer I,  Levitt PR,  Pimenta A,  Zhukareva V;
DR   WPI; 96-455009/45.
DR   N-PSDB; T42080.
PT   DNA encoding limbic system associated membrane protein self binding
PT   domain  - useful in treatment of excessive neural growth in limbic
PT   system, e.g. in animal having epilepsy, Alzheimer's disease or
PT   schizophrenia
PS   Claim 7; Page 40-42; 87pp; English.
CC   The sequences given in W05152-70 represent fragments of limbic system
CC   associated membrane protein (LAMP) self binding domain.  LAMP is a
CC   self-binding, antibody-like cell surface adhesion protein, which causes
CC   the formation of connections between adjacent neurons.  LAMP is bound by
CC   the monoclonal antibody 2G9 and is thought to be involved in the growth
CC   and differentiation of certain neurons.  The protein is highly conserved
CC   and the human and rat sequences differ in only four amino acids.  LAMP
CC   nucleic acids, or soluble LAMP analogues, can be used to treat an
CC   animal with excessive neural growth in the limbic region, i. .e where
CC   the animal has epilepsy, Alzheimer's disease or schizophrenia.  Neural
CC   stem cells transformed with an expression vector comprising one of
CC   these nucleic acids, can be used to treat neuropathologies involving
CC   the limbic system.
SQ   Sequence   338 AA;


  Query Match             5.9%;  Score 85;  DB 1;  Length 338;
  Best Local Similarity   27.7%;  Pred. No. 1.2;
  Matches   36;  Conservative   15;  Mismatches   47;  Indels   32;  Gaps    7;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |     ::   | :|| |   :  |      |     |:| ::
Db     230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy     157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
           || |    |::|       |:| | | :|| |  :|| |     ||           |
Db     286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL----------LA 328

Qy     210 LLKFCTVGFC 219
           || :  |
Db     329 ASLFCLLSKC 338


RESULT  8
W05154
ID   W05154 standard; Protein; 338 AA.
AC   W05154;
DT   22-MAY-1997  (first entry)
DE   Rat LAMP residues 1-332.
KW   limbic system associated membrane protein; LAMP; self binding domain;
KW   antibody-like; cell surface adhesion protein; neuron;
KW   monoclonal antibody; 2G9; growth; differentiation; epilepsy;
KW   Alzheimer's disease; schizophrenia; neural stem cell.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   peptide         1. .28
FT                   /note= "Signal peptide"
FT   protein         29. .338
FT                   /note= "Mature protein"
PN   WO9630052-A1.
PD   03-OCT-1996.
PF   29-MAR-1996; U04397.
PR   31-MAR-1995; US-414657.
PA   (UMDJ-) UMDNJ UNIV NEW JERSEYS HEALTH SCI.
PI   Fischer I,  Levitt PR,  Pimenta A,  Zhukareva V;
DR   WPI; 96-455009/45.
DR   N-PSDB; T42081.
PT   DNA encoding limbic system associated membrane protein self binding
PT   domain  - useful in treatment of excessive neural growth in limbic
PT   system, e.g. in animal having epilepsy, Alzheimer's disease or
PT   schizophrenia
PS   Claim 4; Page 43-45; 87pp; English.
CC   The sequences given in W05152-70 represent fragments of limbic system
CC   associated membrane protein (LAMP) self binding domain.  LAMP is a
CC   self-binding, antibody-like cell surface adhesion protein, which causes
CC   the formation of connections between adjacent neurons.  LAMP is bound by
CC   the monoclonal antibody 2G9 and is thought to be involved in the growth
CC   and differentiation of certain neurons.  The protein is highly conserved
CC   and the human and rat sequences differ in only four amino acids.  LAMP
CC   nucleic acids, or soluble LAMP analogues, can be used to treat an
CC   animal with excessive neural growth in the limbic region, i. .e where
CC   the animal has epilepsy, Alzheimer's disease or schizophrenia.  Neural
CC   stem cells transformed with an expression vector comprising one of
CC   these nucleic acids, can be used to treat neuropathologies involving
CC   the limbic system.
SQ   Sequence   338 AA;


  Query Match             5.9%;  Score 85;  DB 1;  Length 338;
  Best Local Similarity   27.7%;  Pred. No. 1.2;
  Matches   36;  Conservative   15;  Mismatches   47;  Indels   32;  Gaps    7;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |     ::   | :|| |   :  |      |     |:| ::
Db     230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy     157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
           || |    |::|       |:| | | :|| |  :|| |     ||           |
Db     286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL----------LA 328

Qy     210 LLKFCTVGFC 219
           || :  |
Db     329 ASLFCLLSKC 338


RESULT  9
W05156
ID   W05156 standard; Protein; 310 AA.
AC   W05156;
DT   22-MAY-1997  (first entry)
DE   Rat mature LAMP.
KW   limbic system associated membrane protein; LAMP; self binding domain;
KW   antibody-like; cell surface adhesion protein; neuron;
KW   monoclonal antibody; 2G9; growth; differentiation; epilepsy;
```

KW Alzheimer's disease; schizophrenia; neural stem cell.
OS Rattus rattus.
PN WO9630052-A1.
PD 03-OCT-1996.
PF 29-MAR-1996; U04397.
PR 31-MAR-1995; US-414657.
PA (UMDJ-) UMDNJ UNIV NEW JERSEYS HEALTH SCI.
PI Fischer I, Levitt PR, Pimenta A, Zhukareva V;
DR WPI; 96-455009/45.
DR N-PSDB; T42083.
PT DNA encoding limbic system associated membrane protein self binding
PT domain - useful in treatment of excessive neural growth in limbic
PT system, e.g. in animal having epilepsy, Alzheimer's disease or
PT schizophrenia
PS Claim 4; Page 47-49; 87pp; English.
CC The sequences given in W05152-70 represent fragments of limbic system
CC associated membrane protein (LAMP) self binding domain. LAMP is a
CC self-binding, antibody-like cell surface adhesion protein, which causes
CC the formation of connections between adjacent neurons. LAMP is bound by
CC the monoclonal antibody 2G9 and is thought to be involved in the growth
CC and differentiation of certain neurons. The protein is highly conserved
CC and the human and rat sequences differ in only four amino acids. LAMP
CC nucleic acids, or soluble LAMP analogues, can be used to treat an
CC animal with excessive neural growth in the limbic region, i. .e where
CC the animal has epilepsy, Alzheimer's disease or schizophrenia. Neural
CC stem cells transformed with an expression vector comprising one of
CC these nucleic acids, can be used to treat neuropathologies involving
CC the limbic system.
SQ Sequence 310 AA;

Query Match 5.9%; Score 85; DB 1; Length 310;
Best Local Similarity 27.7%; Pred. No. 1;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

```
Qy    101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
          : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db    202 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 257

Qy    157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
          || |    |::|        |:| | | :|| |  :|| |     ||          |
Db    258 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL----------LA 300

Qy    210 LLKFCTVGFC 219
            || :  |
Db    301 ASLFCLLSKC 310
```

RESULT 10
W94361
ID W94361 standard; Protein; 764 AA.
AC W94361;
DT 13-APR-1999 (first entry)
DE Human Factor B analogue #8.
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW stroke; transplant rejection suppression; myocardial infarction.
OS Homo sapiens.
OS Synthetic.
PN US5869615-A.
PD 09-FEB-1999.
PF 03-JAN-1994; 177109.
PR 03-JAN-1994; US-177109.
PA (UNIW) UNIV WASHINGTON.
PI Hourcade DE, Oglesby TJ;
DR WPI; 99-152877/13.
PT Factor B analogues - with modified complement-mediated activity
PS Claim 18; Column -; 53pp; English.
CC The present sequence represents a human Factor B analogue. The present
CC invention describes human Factor B analogues (see W94353 to W94390) that
CC exhibit modified complement-mediated activity in vitro and consist of
CC the sequence given in W94352, of 764 amino acids, with one or more amino
CC acids in a short consensus repeat (SCR) domain or a von Willebrand
CC Factor (vWF) domain replaced by one or more amino acids from a
CC corresponding region of an SCR or vWF domain of a second protein, or
CC in which a protease domain of native Factor B is replaced by a protease
CC domain of C2. Complement protein analogues having altered activity can
CC be used to control activation of the complement system and are so are
CC useful for treating disorders associated with abnormal and/or excessive
CC complement activation e.g. for treating autoimmune diseases, suppressing
CC transplant rejection or reducing tissue damage associated with myocardial
CC infarction and stroke.
CC N.B. The present sequence is not given in the specification but is
CC derived from the sequence in W94352 as specified in the claims.
SQ Sequence 764 AA;

Query Match 5.9%; Score 85; DB 1; Length 764;
Best Local Similarity 24.1%; Pred. No. 3.7;
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

```
Qy     24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV----------- 61
          |: | |       | |:|| | || ||           || | | | :
Db      2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy     61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
            :    | |  ||  :  | |          :|| |  : |            :: |
Db     56 QALEYVCPQGLYPSPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy    111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
           | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db    106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy    167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db    162 AGYCSNPGIPIGTRKV-GSQYRL 183
```

RESULT 11
W94366
ID W94366 standard; Protein; 764 AA.
AC W94366;
DT 13-APR-1999 (first entry)
DE Human Factor B analogue #13.
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW stroke; transplant rejection suppression; myocardial infarction.
OS Homo sapiens.
OS Synthetic.
PN US5869615-A.
PD 09-FEB-1999.
PF 03-JAN-1994; 177109.
PR 03-JAN-1994; US-177109.
PA (UNIW) UNIV WASHINGTON.
PI Hourcade DE, Oglesby TJ;
DR WPI; 99-152877/13.
PT Factor B analogues - with modified complement-mediated activity
PS Claim 18; Column -; 53pp; English.
CC The present sequence represents a human Factor B analogue. The present
CC invention describes human Factor B analogues (see W94353 to W94390) that
CC exhibit modified complement-mediated activity in vitro and consist of
CC the sequence given in W94352, of 764 amino acids, with one or more amino
CC acids in a short consensus repeat (SCR) domain or a von Willebrand
CC Factor (vWF) domain replaced by one or more amino acids from a
CC corresponding region of an SCR or vWF domain of a second protein, or
CC in which a protease domain of native Factor B is replaced by a protease
CC domain of C2. Complement protein analogues having altered activity can
CC be used to control activation of the complement system and are so are
CC useful for treating disorders associated with abnormal and/or excessive
CC complement activation e.g. for treating autoimmune diseases, suppressing
CC transplant rejection or reducing tissue damage associated with myocardial
CC infarction and stroke.
CC N.B. The present sequence is not given in the specification but is
CC derived from the sequence in W94352 as specified in the claims.
SQ Sequence 764 AA;

Query Match 5.8%; Score 84; DB 1; Length 764;

```
  Best Local Similarity   24.1%;  Pred. No. 4.7;
  Matches   49;  Conservative   22;  Mismatches   70;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV---------- 61
           |: | |       | |:|| | || ||           || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :    |||  |   :  | |          :|| |  : |           : ::|
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECKPVRCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


RESULT  12
W94373
ID   W94373 standard; Protein; 764 AA.
AC   W94373;
DT   13-APR-1999  (first entry)
DE   Human Factor B analogue #20.
KW   Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW   stroke; transplant rejection suppression; myocardial infarction.
OS   Homo sapiens.
OS   Synthetic.
PN   US5869615-A.
PD   09-FEB-1999.
PF   03-JAN-1994; 177109.
PR   03-JAN-1994; US-177109.
PA   (UNIW ) UNIV WASHINGTON.
PI   Hourcade DE, Oglesby TJ;
DR   WPI; 99-152877/13.
PT   Factor B analogues - with modified complement-mediated activity
PS   Claim 18; Column -; 53pp; English.
CC   The present sequence represents a human Factor B analogue. The present
CC   invention describes human Factor B analogues (see W94353 to W94390) that
CC   exhibit modified complement-mediated activity in vitro and consist of
CC   the sequence given in W94352, of 764 amino acids, with one or more amino
CC   acids in a short consensus repeat (SCR) domain or a von Willebrand
CC   Factor (vWF) domain replaced by one or more amino acids from a
CC   corresponding region of an SCR or vWF domain of a second protein, or
CC   in which a protease domain of native Factor B is replaced by a protease
CC   domain of C2. Complement protein analogues having altered activity can
CC   be used to control activation of the complement system and are so are
CC   useful for treating disorders associated with abnormal and/or excessive
CC   complement activation e.g. for treating autoimmune diseases, suppressing
CC   transplant rejection or reducing tissue damage associated with myocardial
CC   infarction and stroke.
CC   N.B. The present sequence is not given in the specification but is
CC   derived from the sequence in W94352 as specified in the claims.
SQ   Sequence   764 AA;


  Query Match             5.8%;  Score 84;  DB 1;  Length 764;
  Best Local Similarity   24.6%;  Pred. No. 4.7;
  Matches   50;  Conservative   20;  Mismatches   71;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV---------- 61
           |: | |       | |:|| | || ||           || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :    |||  |   :  | |          :|| |  : |           :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     ||    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWDGETAVCDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


RESULT  13
W94362
ID   W94362 standard; Protein; 763 AA.
AC   W94362;
DT   13-APR-1999  (first entry)
DE   Human Factor B analogue #9.
KW   Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW   stroke; transplant rejection suppression; myocardial infarction.
OS   Homo sapiens.
OS   Synthetic.
PN   US5869615-A.
PD   09-FEB-1999.
PF   03-JAN-1994; 177109.
PR   03-JAN-1994; US-177109.
PA   (UNIW ) UNIV WASHINGTON.
PI   Hourcade DE, Oglesby TJ;
DR   WPI; 99-152877/13.
PT   Factor B analogues - with modified complement-mediated activity
PS   Claim 18; Column -; 53pp; English.
CC   The present sequence represents a human Factor B analogue. The present
CC   invention describes human Factor B analogues (see W94353 to W94390) that
CC   exhibit modified complement-mediated activity in vitro and consist of
CC   the sequence given in W94352, of 764 amino acids, with one or more amino
CC   acids in a short consensus repeat (SCR) domain or a von Willebrand
CC   Factor (vWF) domain replaced by one or more amino acids from a
CC   corresponding region of an SCR or vWF domain of a second protein, or
CC   in which a protease domain of native Factor B is replaced by a protease
CC   domain of C2. Complement protein analogues having altered activity can
CC   be used to control activation of the complement system and are so are
CC   useful for treating disorders associated with abnormal and/or excessive
CC   complement activation e.g. for treating autoimmune diseases, suppressing
CC   transplant rejection or reducing tissue damage associated with myocardial
CC   infarction and stroke.
CC   N.B. The present sequence is not given in the specification but is
CC   derived from the sequence in W94352 as specified in the claims.
SQ   Sequence   763 AA;


  Query Match             5.8%;  Score 83.5;  DB 1;  Length 763;
  Best Local Similarity   24.6%;  Pred. No. 5.3;
  Matches   49;  Conservative   24;  Mismatches   71;  Indels   55;  Gaps   11;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAV-----T 78
           |: | |       | |:|| | || ||       |: :  : | | |  : | |     :
Db       2 GSNLSP------QLCLMPFILGLLSGG-------VTTTPWSLAQPQGSCSLEGVEIKGGS 48

Qy      79 ARLV---GVLWFVSVT---TGPWGAVATSAGGEESLKCEDLKV---------------- 116
            ||:     | :|  :     |    :  | |   :|| :| |
Db      49 FRLLQEGQALEYVCPSGFYPYPASRLCRSTGSWSTLKTQDQKTVRKAECRAIHCPRPHDF 108

Qy     116 --GQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNEVGFF 170
             |:|  : |  | : :   :|   ||   |     | ||: :     :|    |  |:
Db     109 ENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNGAGYC 164

Qy     171 K----PISCRNVNGYSYKV 185
               ||  | | |  |::
Db     165 SNPGIPIGTRKV-GSQYRL 182


RESULT  14
W94383
ID   W94383 standard; Protein; 764 AA.
AC   W94383;
DT   13-APR-1999  (first entry)
DE   Human Factor B analogue #30.
KW   Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW   stroke; transplant rejection suppression; myocardial infarction.
```

```
OS   Homo sapiens.
OS   Synthetic.
PN   US5869615-A.
PD   09-FEB-1999.
PF   03-JAN-1994; 177109.
PR   03-JAN-1994; US-177109.
PA   (UNIW ) UNIV WASHINGTON.
PI   Hourcade DE, Oglesby TJ;
DR   WPI; 99-152877/13.
PT   Factor B analogues - with modified complement-mediated activity
PS   Claim 18; Column -; 53pp; English.
CC   The present sequence represents a human Factor B analogue. The present
CC   invention describes human Factor B analogues (see W94353 to W94390) that
CC   exhibit modified complement-mediated activity in vitro and consist of
CC   the sequence given in W94352, of 764 amino acids, with one or more amino
CC   acids in a short consensus repeat (SCR) domain or a von Willebrand
CC   Factor (vWF) domain replaced by one or more amino acids from a
CC   corresponding region of an SCR or vWF domain of a second protein, or
CC   in which a protease domain of native Factor B is replaced by a protease
CC   domain of C2. Complement protein analogues having altered activity can
CC   be used to control activation of the complement system and are so are
CC   useful for treating disorders associated with abnormal and/or excessive
CC   complement activation e.g. for treating autoimmune diseases, suppressing
CC   transplant rejection or reducing tissue damage associated with myocardial
CC   infarction and stroke.
CC   N.B. The present sequence is not given in the specification but is
CC   derived from the sequence in W94352 as specified in the claims.
SQ   Sequence   764 AA;

  Query Match             5.8%;  Score 83;  DB 1;  Length 764;
  Best Local Similarity   24.1%;  Pred. No. 6;
  Matches   49;  Conservative   21;  Mismatches   71;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV----------- 61
           |: | |       | |:|| | || ||            || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :    |||  |   :  | |          :|| |  : |            :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
            |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


RESULT  15
W94384
ID   W94384 standard; Protein; 764 AA.
AC   W94384;
DT   13-APR-1999  (first entry)
DE   Human Factor B analogue #31.
KW   Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW   stroke; transplant rejection suppression; myocardial infarction.
OS   Homo sapiens.
OS   Synthetic.
PN   US5869615-A.
PD   09-FEB-1999.
PF   03-JAN-1994; 177109.
PR   03-JAN-1994; US-177109.
PA   (UNIW ) UNIV WASHINGTON.
PI   Hourcade DE, Oglesby TJ;
DR   WPI; 99-152877/13.
PT   Factor B analogues - with modified complement-mediated activity
PS   Claim 18; Column -; 53pp; English.
CC   The present sequence represents a human Factor B analogue. The present
CC   invention describes human Factor B analogues (see W94353 to W94390) that
CC   exhibit modified complement-mediated activity in vitro and consist of
CC   the sequence given in W94352, of 764 amino acids, with one or more amino
CC   acids in a short consensus repeat (SCR) domain or a von Willebrand
CC   Factor (vWF) domain replaced by one or more amino acids from a
CC   corresponding region of an SCR or vWF domain of a second protein, or
CC   in which a protease domain of native Factor B is replaced by a protease
CC   domain of C2. Complement protein analogues having altered activity can
CC   be used to control activation of the complement system and are so are
CC   useful for treating disorders associated with abnormal and/or excessive
CC   complement activation e.g. for treating autoimmune diseases, suppressing
CC   transplant rejection or reducing tissue damage associated with myocardial
CC   infarction and stroke.
CC   N.B. The present sequence is not given in the specification but is
CC   derived from the sequence in W94352 as specified in the claims.
SQ   Sequence   764 AA;

  Query Match             5.8%;  Score 83;  DB 1;  Length 764;
  Best Local Similarity   24.1%;  Pred. No. 6;
  Matches   49;  Conservative   21;  Mismatches   71;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV----------- 61
           |: | |       | |:|| | || ||            || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLAQPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :    |||  |   :  | |          :|| |  : |            :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
            |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183


Search completed: February 11, 2000, 04:17:02
Job time: 80 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 05:16:36 ; Search time 32.32 Seconds
(without alignments)
248.566 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 269
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: OLIGO

Searched: 82229 seqs, 29864866 residues

base : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 82229

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	3.0	1520	1	ABL_DROME	P00522 drosophila
2	8	3.0	1174	1	CIKE_DROME	Q02280 drosophila
3	8	3.0	118	1	CRYP_CRYPA	P52753 cryphonectr
4	8	3.0	473	1	HN3A_HUMAN	P55317 homo sapien
5	8	3.0	358	1	KKIA_HUMAN	Q00532 homo sapien
6	8	3.0	613	1	XPCT_HUMAN	P36021 homo sapien
7	8	3.0	427	1	YEAH_ECOLI	P76235 escherichia
8	8	3.0	753	1	YJ05_CAEEL	O01975 caenorhabdi
9	7	2.6	1039	1	AG43_ECOLI	P39180 escherichia
10	7	2.6	1959	1	AGRI_RAT	P25304 rattus norv
11	7	2.6	882	1	AREA_ASPNG	O13412 aspergillus
12	7	2.6	550	1	CCF_DROME	P41046 drosophila
13	7	2.6	581	1	CEA2_ECOLI	P04419 escherichia
14	7	2.6	551	1	CEA3_ECOLI	P00646 escherichia
15	7	2.6	551	1	CEA6_ECOLI	P17999 escherichia
16	7	2.6	582	1	CEA9_ECOLI	P09883 escherichia
17	7	2.6	510	1	CEAB_ECOLI	P05819 escherichia
18	7	2.6	663	1	CNG2_BOVIN	Q03041 bos taurus
19	7	2.6	424	1	COAA_BPFD	P03661 bacteriopha
20	7	2.6	424	1	COAA_BPM13	P03662 bacteriopha
21	7	2.6	419	1	CYB_RHOVI	P81378 rhodopseudo
22	7	2.6	2038	1	FSH_DROME	P13709 drosophila
23	7	2.6	907	1	GLR1_MOUSE	P23818 mus musculu
24	7	2.6	572	1	GPC5_HUMAN	P78333 homo sapien
25	7	2.6	183	1	GRP2_ORYSA	P29834 oryza sativ
26	7	2.6	173	1	IPYR_HELPY	P56153 helicobacte
27	7	2.6	575	1	ITR1_SCHPO	Q10286 schizosacch
28	7	2.6	622	1	K1CI_HUMAN	P35527 homo sapien
29	7	2.6	166	1	K2C5_BOVIN	P04262 bos taurus
30	7	2.6	796	1	KF3C_MOUSE	O35066 mus musculu
31	7	2.6	638	1	LIK2_RAT	P53670 rattus norv
32	7	2.6	1192	1	LMG2_MOUSE	Q61092 mus musculu
33	7	2.6	534	1	ML51_HUMAN	O15234 homo sapien
34	7	2.6	399	1	MPK4_HUMAN	P45985 homo sapien
35	7	2.6	577	1	MPPA_NEUCR	P23955 neurospora
36	7	2.6	482	1	MPPA_YEAST	P11914 saccharomyc
37	7	2.6	375	1	PER_DROCP	P91607 drosophila
38	7	2.6	375	1	PER_DROSC	P91705 drosophila
39	7	2.6	435	1	PGLX_ASPTU	Q00293 aspergillus
40	7	2.6	623	1	PNT1_DROME	P51022 drosophila
41	7	2.6	718	1	PNT2_DROME	P51023 drosophila
42	7	2.6	3080	1	POLG_ZYMVC	P18479 z genome po
43	7	2.6	163	1	PTFB_BACSU	P26380 bacillus su
44	7	2.6	307	1	RB33_CAEEL	Q20365 caenorhabdi
45	7	2.6	979	1	RFX1_HUMAN	P22670 homo sapien

ALIGNMENTS

```
RESULT   1
ABL_DROME
ID   ABL_DROME      STANDARD;      PRT;  1520 AA.
AC   P00522;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   TYROSINE-PROTEIN KINASE DASH/ABL (EC 2.7.1.112).
GN   DASH OR ABL-1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88174728.
RA   HENKEMEYER M.J., BENNETT R.L., GERTLER F.B., HOFFMANN F.M.;
RT   "DNA sequence, structure, and tyrosine kinase activity of the
RT   Drosophila melanogaster Abelson proto-oncogene homolog.";
RL   Mol. Cell. Biol. 8:843-853(1988).
RN   [2]
RP   SEQUENCE OF 374-648 FROM N.A.
RX   MEDLINE; 84082064.
RA   HOFFMANN F.M., FRESCO L.D., HOFFMAN-FALK H., SHILO B.-Z.;
RT   "Nucleotide sequences of the Drosophila src and abl homologs:
RT   conservation and variability in the src family oncogenes.";
RL   Cell 35:393-401(1983).
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. BELONGS TO THE ABL SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19692; AAA28934.1; -.
DR   EMBL; M19690; AAA28934.1; JOINED.
DR   EMBL; M19691; AAA28934.1; JOINED.
DR   EMBL; K01042; AAA28443.1; -.
DR   PIR; A28128; TVFFA.
DR   HSSP; P00519; 1AB2.
DR   FLYBASE; FBgn0000017; Abl.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   SH2 domain; SH3 domain.
FT   DOMAIN      204    265       SH3.
FT   DOMAIN      271    363       SH2.
```

```
FT   DOMAIN      388    644       PROTEIN KINASE.
FT   NP_BIND     394    402       ATP (BY SIMILARITY).
FT   BINDING     417    417       ATP (BY SIMILARITY).
FT   ACT_SITE    509    509       BY SIMILARITY.
FT   MOD_RES     539    539       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT    374    377       LSPE -> ASAQ (IN REF. 2).
FT   CONFLICT    645    648       ESSI -> VGDV (IN REF. 2).
SQ   SEQUENCE   1520 AA;  161836 MW;  AE831F46 CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 1520;
  Best Local Similarity   100.0%;  Pred. No. 9.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db      78 GGGGSGSG 85


RESULT   2
CIKE_DROME
ID   CIKE_DROME     STANDARD;      PRT;  1174 AA.
AC   Q02280;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   POTASSIUM CHANNEL PROTEIN EAG.
GN   EAG.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91262635.
RA   WARMKE J., DRYSDALE R., GANETZKY B.;
RT   "A distinct potassium channel polypeptide encoded by the Drosophila
RT   eag locus.";
RL   Science 252:1560-1562(1991).
CC   -!- FUNCTION: PROTEIN EAG IS MOST PROBABLY A STRUCTURAL COMPONENT OF
CC       THE POTASSIUM CHANNEL AND MEDIATES THE POTASSIUM PERMEABILITY OF
CC       MEMBRANES.
CC   -!- SUBUNIT: DIMER (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND
CC       IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
C        EVERY THIRD POSITION.
C    -!- MISCELLANEOUS: THE SEGMENT H5 IS THOUGHT TO LINE THE CHANNEL PORE.
C    -!- SIMILARITY: TO THE MEMBERS OF THE POTASSIUM CHANNEL PROTEINS
CC       OF THE SH SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61157; AAA28495.1; -.
DR   PIR; A40853; A40853.
DR   FLYBASE; FBgn0000535; eag.
DR   PFAM; PF00027; cNMP_binding; 1.
DR   PFAM; PF00914; CNG_membrane; 1.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Glycoprotein; Phosphorylation; Repeat.
FT   REPEAT       18     91       PAS-1.
FT   DOMAIN      116    156       PAC MOTIF.
FT   TRANSMEM    227    246       SEGMENT S1 (POTENTIAL).
FT   TRANSMEM    269    291       SEGMENT S2 (POTENTIAL).
FT   TRANSMEM    314    335       SEGMENT S3 (POTENTIAL).
FT   TRANSMEM    343    369       SEGMENT S4 (POTENTIAL).
FT   TRANSMEM    372    393       SEGMENT S5 (POTENTIAL).
FT   TRANSMEM    442    467       SEGMENT H5 (POTENTIAL).
FT   TRANSMEM    471    493       SEGMENT S6 (POTENTIAL).
FT   CARBOHYD    262    262       POTENTIAL.
FT   CARBOHYD    412    412       POTENTIAL.
FT   CARBOHYD    424    424       POTENTIAL.
SQ   SEQUENCE   1174 AA;  126236 MW;  EF855F66 CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 7.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db    1125 GGGGSGSG 1132


RESULT   3
CRYP_CRYPA
ID   CRYP_CRYPA     STANDARD;      PRT;   118 AA.
AC   P52753;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   CRYPARIN PRECURSOR.
GN   CRP.
OS   Cryphonectria parasitica (Chesnut blight fungus) (Endothia
OS   parasitica).
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC   Diaporthales; Valsaceae; Cryphonectria.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-73.
RC   STRAIN=155/2;
RX   MEDLINE; 94156182.
RA   ZHANG L., VILLALON D., SUN Y., KAZMIERCZAK P., VAN ALFEN N.K.;
RT   "Virus-associated down-regulation of the gene encoding cryparin, an
RT   abundant cell-surface protein from the chestnut blight fungus,
RT   Cryphonectria parasitica.";
RL   Gene 139:59-64(1994).
CC   -!- FUNCTION: CONTRIBUTES TO SURFACE HYDROPHOBICITY, WHICH IS
CC       IMPORTANT FOR PROCESSES SUCH AS ASSOCIATION OF HYPHAE IN
CC       REPRODUCTIVE STRUCTURES, DISPERSAL OF AERIAL SPORES AND ADHESION
CC       OF PATHOGENS TO HOST STRUCTURES. PRODUCED ABUNDANTLY, EXCEPT IN
CC       THE DS-RNA VIRUS-INFECTED STRAINS, WHERE THE EXPRESSION IS MUCH
CC       REDUCED.
CC   -!- SUBCELLULAR LOCATION: CELL WALL OF AERIAL HYPHAE AND SPORULATION
CC       STRUCTURES. ABUNDANTLY SECRETED.
CC   -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED ON DAY 2 AND 3 AFTER
CC       INOCULATION, A TIME WHEN THE FUNGUS IS IN A RAPID PHASE OF GROWTH.
CC       AFTER A STATIONARY PHASE ON DAY 4, THE EXPRESSION DECREASES.
CC   -!- SIMILARITY: BELONGS TO THE CERATO-ULMIN HYDROPHOBIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09559; AAA19638.1; -.
KW   Cell wall; Signal; Repeat.
FT   SIGNAL        1     22
FT   CHAIN        23    118       CRYPARIN.
FT   DOMAIN       23     32       POLY-GLY.
FT   REPEAT       29     42       7 X 2 AA TANDEM REPEAT OF S-G.
FT   REPEAT       29     30       1.
FT   REPEAT       31     32       2.
FT   REPEAT       33     34       3.
FT   REPEAT       35     36       4.
FT   REPEAT       37     38       5.
FT   REPEAT       39     40       6.
```

```
FT   REPEAT       41     42       7.
SQ   SEQUENCE   118 AA;  11387 MW;  5B2C48A6 CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 GGGGSGSG 55
          ||||||||
Db     25 GGGGSGSG 32


RESULT   4
HN3A_HUMAN
ID   HN3A_HUMAN     STANDARD;      PRT;   473 AA.
AC   P55317;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HEPATOCYTE NUCLEAR FACTOR 3-ALPHA (HNF-3A).
GN   HNF3A OR TCF3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 96254057.
RA   BINGLE C.D., GOWAN S.;
RT   "Molecular cloning of the forkhead transcription factor HNF-3 alpha
RT   from a human pulmonary adenocarcinoma cell line.";
RL   Biochim. Biophys. Acta 1307:17-20(1996).
CC   -!- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES
CC       SUCH AS AFP, ALBUMIN, TYROSINE AMINOTRANSFERASE, PEPCK, ETC.
CC       INTERACTS WITH THE CIS-ACTING REGULATORY REGIONS OF THESE GENES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39840; AAB06493.1; -.
DR   MIM; 602294; -.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PFAM; PF00250; Fork_head; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT   DNA_BIND    170    261       FORK-HEAD.
SQ   SEQUENCE   473 AA;  49321 MW;  AE4F7931 CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 GGGGSGSG 55
          ||||||||
Db    277 GGGGSGSG 284


RESULT   5
KKIA_HUMAN
ID   KKIA_HUMAN     STANDARD;      PRT;   358 AA.
AC   Q00532;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   SERINE/THREONINE-PROTEIN KINASE KKIALRE (EC 2.7.1.-).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92347325.
RA   MEYERSON M., ENDERS G.H., WU C.-L., SU L.-K., GORKA C., NELSON C.,
RA   HARLOW E., TSAI L.-H.;
RT   "A family of human cdc2-related protein kinases.";
RL   EMBO J. 11:2909-2917(1992).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       CDC2/CDKX SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66358; CAA47002.1; -.
DR   EMBL; X66359; CAA47002.1; JOINED.
DR   PIR; S22744; S22744.
DR   PIR; S22745; S22745.
DR   PIR; S23383; S23383.
DR   HSSP; P24941; 1AQ1.
DR   MIM; 603441; -.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN        5    288       PROTEIN KINASE.
FT   NP_BIND      11     19       ATP (BY SIMILARITY).
FT   BINDING      34     34       ATP (BY SIMILARITY).
FT   ACT_SITE    127    127       BY SIMILARITY.
SQ   SEQUENCE   358 AA;  41834 MW;  87B81E2E CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    205 YPALGLLK 212
          ||||||||
Db    259 YPALGLLK 266


RESULT   6
XPCT_HUMAN
ID   XPCT_HUMAN     STANDARD;      PRT;   613 AA.
AC   P36021;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   X-LINKED PEST-CONTAINING TRANSPORTER.
GN   SLC16A2 OR XPCT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95072579.
RA   LAFRENIERE R.G., CARREL L., WILLARD H.F.;
RT   "A novel transmembrane transporter encoded by the XPCT gene in
RT   Xq13.2.";
RL   Hum. Mol. Genet. 3:1133-1140(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MONOCARBOXYLATE TRANSPORTER FAMILY.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U05321; AAB60375.1; -.
DR   EMBL; U05316; AAB60375.1; JOINED.
DR   EMBL; U05317; AAB60375.1; JOINED.
DR   EMBL; U05318; AAB60375.1; JOINED.
DR   EMBL; U05319; AAB60375.1; JOINED.
DR   EMBL; U05320; AAB60375.1; JOINED.
DR   EMBL; U05315; AAB60374.1; -.
DR   MIM; 300095; -.
DR   PFAM; PF01587; MCT; 1.
KW   Transport; Transmembrane; Glycoprotein.
FT   DOMAIN       1     70       GLY/SER-RICH.
FT   TRANSMEM   171    191       POTENTIAL.
FT   TRANSMEM   218    238       POTENTIAL.
FT   TRANSMEM   246    266       POTENTIAL.
FT   TRANSMEM   275    295       POTENTIAL.
FT   TRANSMEM   304    324       POTENTIAL.
FT   TRANSMEM   333    353       POTENTIAL.
FT   TRANSMEM   397    417       POTENTIAL.
FT   TRANSMEM   431    451       POTENTIAL.
FT   TRANSMEM   461    481       POTENTIAL.
FT   TRANSMEM   484    504       POTENTIAL.
FT   TRANSMEM   522    542       POTENTIAL.
FT   TRANSMEM   552    572       POTENTIAL.
SQ   SEQUENCE   613 AA;  66400 MW;  D5191340 CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 4.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db      34 GGGGSGSG 41


RESULT   7
YEAH_ECOLI
ID   YEAH_ECOLI     STANDARD;      PRT;   427 AA.
AC   P76235; O07964;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL 49.4 KD PROTEIN IN GAPA-RND INTERGENIC REGION.
GN   YEAH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97251358.
RA   ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RA   KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,
RA   MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,
RA   NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA   SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,
RA   YAMAMOTO Y., HORIUCHI T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
CC   -!- SIMILARITY: TO B.SUBTILIS YHBH.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000273; AAC74854.1; -.
DR   EMBL; D90822; CAB21509.1; -.
DR   ECOGENE; EG13494; YEAH.
KW   Hypothetical protein.
SQ   SEQUENCE   427 AA;  49392 MW;  42F92C9E CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db      94 GGGGSGSG 101


RESULT   8
YJ05_CAEEL
ID   YJ05_CAEEL     STANDARD;      PRT;   753 AA.
AC   O01975;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 85.1 KD NUCLEASE C41D11.5 (EC 3.1.30.-).
GN   C41D11.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GATTUNG S., MAGGI L.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE DNA/RNA NON-SPECIFIC ENDONUCLEASES
CC       FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF003740; AAC48141.1; -.
DR   WORMPEP; C41D11.5; CE08662.
DR   PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR   PFAM; PF01223; Endonuclease; 1.
KW   Hypothetical protein; Hydrolase; Nuclease; Endonuclease.
FT   ACT_SITE    593    593       BY SIMILARITY.
SQ   SEQUENCE   753 AA;  85068 MW;  9DF3803A CRC32;


  Query Match             3.0%;  Score 8;  DB 1;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     197 GADRFYLG 204
```

```
          !!!!!!!!
Db    398 GADRFYLG 405


RESULT   9
AG43_ECOLI
ID   AG43_ECOLI     STANDARD;      PRT;  1039 AA.
AC   P39180; P76360; P75614; P97241; Q46771;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ANTIGEN 43 PRECURSOR (AG43) (FLUFFING PROTEIN).
GN   FLU.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97251358.
RA   ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RA   KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,
RA   MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,
RA   NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA   SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,
RA   YAMAMOTO Y., HORIUCHI T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ML 308-225;
RA   HENDERSON I.R., OWEN P.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   PARTIAL SEQUENCE.
RC   STRAIN=ML 308-225;
RX   MEDLINE; 89291704.
RA   CAFFREY P., OWEN P.;
RT   "Purification and N-terminal sequence of the alpha subunit of antigen
RT   43, a unique protein complex associated with the outer membrane of
RT   Escherichia coli.";
RL   J. Bacteriol. 171:3634-3640(1989).
RN   [5]
RP   SEQUENCE OF 53-63.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE; 97443975.
RA   LINK A.J., ROBISON K., CHURCH G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [6]
RP   GENE NAME.
RX   MEDLINE; 97257509.
RA   HENDERSON I.R., MEEHAN M., OWEN P.;
RT   "Antigen 43, a phase-variable bipartite outer membrane protein,
RT   determines colony morphology and autoaggregation in Escherichia coli
RT   K-12.";
RL   FEMS Microbiol. Lett. 149:115-120(1997).
CC   -!- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION.
CC   -!- SUBUNIT: CONSIST OF TWO SUBUNITS; ALPHA AND BETA.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC   -!- SIMILARITY: TO ADHESIN AIDI-I AND TO TO BORDETELLA PERTACTIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000291; AAC75061.1; ALT_INIT.
DR   EMBL; D90838; CAB21751.1; ALT_INIT.
DR   EMBL; D90839; CAB21756.1; ALT_INIT.
DR   EMBL; U24429; AAB47869.1; -.
DR   ECOGENE; EG12686; FLU.
KW   Outer membrane; Signal.
FT   SIGNAL        1     52
FT   CHAIN        53    551       ANTIGEN 43 ALPHA CHAIN.
FT   CHAIN       552   1039       ANTIGEN 43 BETA CHAIN.
FT   VARIANT       2      2       K -> N (IN STRAIN ML 308-225).
FT   VARIANT      41     42       SL -> FF (IN STRAIN ML 308-225).
FT   VARIANT      46     46       T -> K (IN STRAIN ML 308-225).
FT   VARIANT     157    157       W -> L (IN STRAIN ML 308-225).
FT   VARIANT     188    188       V -> F (IN STRAIN ML 308-225).
FT   VARIANT     303    305       ATN -> STI (IN STRAIN ML 308-225).
FT   VARIANT     320    320       A -> T (IN STRAIN ML 308-225).
FT   VARIANT     372    372       N -> Q (IN STRAIN ML 308-225).
FT   VARIANT     493    493       E -> V (IN STRAIN ML 308-225).
FT   VARIANT     497    497       S -> N (IN STRAIN ML 308-225).
FT   VARIANT     585    585       H -> Y (IN STRAIN ML 308-225).
FT   VARIANT     709    709       E -> K (IN STRAIN ML 308-225).
FT   VARIANT     721    721       M -> T (IN STRAIN ML 308-225).
FT   VARIANT     751    753       GHL -> SHF (IN STRAIN ML 308-225).
FT   VARIANT     803    803       S -> P (IN STRAIN ML 308-225).
FT   VARIANT     815    815       A -> V (IN STRAIN ML 308-225).
FT   VARIANT     824    824       C -> S (IN STRAIN ML 308-225).
FT   VARIANT     829    835       LNLVHTS -> MNLIYNA (IN STRAIN ML 308-
FT                                225).
FT   VARIANT     845    847       QGT -> LGA (IN STRAIN ML 308-225).
FT   VARIANT     855    855       S -> T (IN STRAIN ML 308-225).
FT   VARIANT     888    888       Q -> L (IN STRAIN ML 308-225).
FT   VARIANT    1025   1025       S -> I (IN STRAIN ML 308-225).
FT   CONFLICT     61     63       ETV -> TTT (IN REF. 5).
FT   CONFLICT    824    824       C -> S (IN REF. 2).
SQ   SEQUENCE   1039 AA;  106841 MW;  49DC5F84 CRC32;


  Query Match             2.6%;  Score 7;  DB 1;  Length 1039;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    185 VAVALSL 191
          |||||||
Db     36 VAVALSL 42


RESULT  10
AGRI_RAT
ID   AGRI_RAT       STANDARD;      PRT;  1959 AA.
AC   P25304; Q63034;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   AGRIN PRECURSOR.
GN   AGRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC   TISSUE=EMBRYONIC SPINAL CORD;
RX   MEDLINE; 91222570.
RA   RUPP F., PAYAN D.G., MAGILL-SOLC C., COWAN D.M., SCHELLER R.H.;
RT   "Structure and expression of a rat agrin.";
```

```
RL   Neuron 6:811-823(1991).
RN   [2]
RP   SEQUENCE OF 1777-1801 FROM N.A.
RX   MEDLINE; 92407628.
RA   RUPP F., OEZCELIK T., LINIAL M., PETERSON K., FRANCKE U., SCHELLER R.;
RT   "Structure and chromosomal localization of the mammalian agrin gene.";
RL   J. Neurosci. 12:3535-3544(1992).
CC   -!- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
CC       AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
CC       ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
CC   -!- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
CC       JUNCTION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST FIVE DIFFERENT FORMS ARISE BY
CC       ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR
CC       CLUSTERING ACTIVITY.
CC   -!- TISSUE SPECIFICITY: EMBRYONIC NERVOUS SYSTEM AND MUSCLE.
CC   -!- DEVELOPMENTAL STAGE: MORE ABUNDANT EARLY IN DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 8 KAZAL-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-18 OR MET-24 IS THE
CC       INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64780; AAA40703.1; -.
DR   EMBL; M64780; AAA40702.1; ALT_INIT.
DR   EMBL; S44194; AAB23326.1; -.
DR   PIR; JH0399; AGRT.
DR   HSSP; P00740; 1IXA.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PFAM; PF00050; kazal; 9.
DR   PFAM; PF00053; laminin_EGF; 2.
DR   PFAM; PF00054; laminin_G; 3.
DR   PFAM; PF01390; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain.
FT   SIGNAL       1     29       POTENTIAL.
FT   CHAIN       30   1959       AGRIN.
FT   DOMAIN      65    137       KAZAL-LIKE 1.
FT   DOMAIN     141    212       KAZAL-LIKE 2.
FT   DOMAIN     213    284       KAZAL-LIKE 3.
FT   DOMAIN     287    356       KAZAL-LIKE 4.
FT   DOMAIN     361    429       KAZAL-LIKE 5.
FT   DOMAIN     430    494       KAZAL-LIKE 6.
FT   DOMAIN     495    559       KAZAL-LIKE 7.
FT   DOMAIN     563    645       KAZAL-LIKE 8.
FT   DOMAIN     688    741       LAMININ EGF-LIKE 1.
FT   DOMAIN     742    788       LAMININ EGF-LIKE 2.
FT   DOMAIN     794    864       KAZAL-LIKE.
FT   DOMAIN    1220   1258       EGF-LIKE 1.
FT   DOMAIN    1440   1477       EGF-LIKE 2.
FT   DOMAIN    1479   1516       EGF-LIKE 3.
FT   DOMAIN    1709   1748       EGF-LIKE 4.
FT   DOMAIN     869    992       SER/THR-RICH.
FT   DOMAIN    1147   1215       SER/THR-RICH.
FT   DISULFID    97    116       POTENTIAL.
FT   DISULFID   105    137       POTENTIAL.
FT   DISULFID   171    191       POTENTIAL.
FT   DISULFID   180    212       POTENTIAL.
FT   DISULFID   244    263       POTENTIAL.
FT   DISULFID   252    284       POTENTIAL.
FT   DISULFID   316    335       POTENTIAL.
FT   DISULFID   324    356       POTENTIAL.
FT   DISULFID   389    408       POTENTIAL.
FT   DISULFID   397    429       POTENTIAL.
FT   DISULFID   454    473       POTENTIAL.
FT   DISULFID   462    494       POTENTIAL.
FT   DISULFID   518    538       POTENTIAL.
FT   DISULFID   527    559       POTENTIAL.
FT   DISULFID   604    624       POTENTIAL.
FT   DISULFID   613    645       POTENTIAL.
FT   DISULFID   688    700       BY SIMILARITY.
FT   DISULFID   690    707       BY SIMILARITY.
FT   DISULFID   709    718       BY SIMILARITY.
FT   DISULFID   721    739       BY SIMILARITY.
FT   DISULFID   742    754       BY SIMILARITY.
FT   DISULFID   744    761       BY SIMILARITY.
FT   DISULFID   763    772       BY SIMILARITY.
FT   DISULFID   775    786       BY SIMILARITY.
FT   DISULFID   823    843       POTENTIAL.
FT   DISULFID   832    864       POTENTIAL.
FT   DISULFID  1224   1235       BY SIMILARITY.
FT   DISULFID  1229   1246       BY SIMILARITY.
FT   DISULFID  1248   1257       BY SIMILARITY.
FT   DISULFID  1444   1455       POTENTIAL.
FT   DISULFID  1449   1465       POTENTIAL.
FT   DISULFID  1467   1476       POTENTIAL.
FT   DISULFID  1483   1494       BY SIMILARITY.
FT   DISULFID  1488   1504       BY SIMILARITY.
FT   DISULFID  1506   1515       BY SIMILARITY.
FT   DISULFID  1713   1727       BY SIMILARITY.
FT   DISULFID  1721   1736       BY SIMILARITY.
FT   DISULFID  1738   1747       BY SIMILARITY.
FT   CARBOHYD   145    145       POTENTIAL.
FT   CARBOHYD   672    672       POTENTIAL.
FT   CARBOHYD   827    827       POTENTIAL.
FT   CARBOHYD   957    957       POTENTIAL.
FT   VARSPLIC  1144   1152       MISSING (IN ISOFORM 2).
FT   VARSPLIC  1780   1798       MISSING (IN ISOFORM 3).
FT   VARSPLIC  1788   1798       MISSING (IN ISOFORM 4).
FT   VARSPLIC  1780   1787       MISSING (IN ISOFORM 5).
FT   VARIANT    314    314       V -> VTCD (IN A VARIANT).
SQ   SEQUENCE   1959 AA;  208645 MW;  6B6E022E CRC32;


  Query Match             2.6%;  Score 7;  DB 1;  Length 1959;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      50 GGSGSGE 56
           |||||||
Db     567 GGSGSGE 573


RESULT  11
AREA_ASPNG
ID   AREA_ASPNG     STANDARD;      PRT;   882 AA.
AC   O13412;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NITROGEN REGULATORY PROTEIN AREA.
GN   AREA.
OS   Aspergillus niger.
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N402;
RA   MACCABE A.P., VANHANEN S.A.S., SOLLEWIJN GELPKE M.,
RA   VAN DE VONDERVOORT P., ARST H.N., VISSER J.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC       REGULATORY GENE OF NITROGEN METABOLITE REPRESSION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
CC       HIGH, TO OTHER FUNGAL NITROGEN REGULATORY PROTEINS.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X81998; CAA57524.1; -.
DR   HSSP; P17429; 5GAT.
DR   PROSITE; PS00344; GATA_ZN_FINGER; 1.
DR   PFAM; PF00320; GATA; 1.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW   Nuclear protein; Nitrate assimilation.
FT   ZN_FING     676    700       GATA-TYPE.
SQ   SEQUENCE   882 AA;  94518 MW;  54DEA816 CRC32;


  Query Match             2.6%;  Score 7;  DB 1;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      47 LGGGGSG 53
           |||||||
Db       6 LGGGGSG 12


RESULT  12
CCF_DROME
ID   CCF_DROME      STANDARD;      PRT;   550 AA.
AC   P41046; Q24271;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CENTROSOMAL AND CHROMOSOMAL FACTOR (CCF) (CHROMOCENTROSOMIN).
GN   CORTO.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 98130597.
RA   KODJABACHIAN L., DELAAGE M., MAUREL C., MIASSOD R., JACQ B.,
RA   ROSSET R.;
RT   "Mutations in ccf, a novel Drosophila gene encoding a chromosomal
RT   factor, affect progression through mitosis and interact with Pc-G
RT   mutations.";
RL   EMBO J. 17:1063-1075(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   TAKAMURA C., IMAMURA Y., TAIRA T., IGUCHI-ARIGA S., ARIGA H.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR PROPER CONDENSATION OF MITOTIC CHROMOSOMES
CC       AND PROGRESSION THROUGH MITOSIS. IS AN ESSENTIAL GENE. BINDS TO
CC       SPECIFIC POLYTENE CHROMOSOME SITES, MANY OF WHICH ARE SHARED WITH
CC       THE POSTERIOR SEX COMBS PROTEIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR AND CENTROSOMAL.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING OOGENESIS, EMBRYONIC AND
CC       LARVAL STAGES.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS
CC       FRAMESHIFTS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U35074; AAC13917.1; -.
DR   EMBL; D43795; BAA07853.1; ALT_FRAME.
DR   HSSP; P04002; 1WFA.
DR   FLYBASE; FBgn0010313; corto.
KW   DNA-binding; Nuclear protein; Mitosis; Coiled coil.
FT   DOMAIN       26     35       POLY-GLN.
FT   DOMAIN       46     50       POLY-GLN.
FT   DOMAIN       66     72       POLY-GLN.
FT   DOMAIN       84     89       POLY-SER.
FT   DOMAIN      110    126       POLY-GLN.
FT   DOMAIN      204    210       POLY-SER.
FT   DOMAIN      215    220       POLY-GLY.
FT   DOMAIN      236    275       COILED COIL (POTENTIAL).
FT   DOMAIN      243    250       POLY-GLN.
FT   DOMAIN      253    258       POLY-GLN.
FT   DOMAIN      285    297       POLY-GLN.
FT   DOMAIN      322    329       POLY-ALA.
FT   DOMAIN      332    336       POLY-ALA.
FT   DOMAIN      396    410       POLY-GLN.
FT   DOMAIN      414    417       POLY-PRO.
FT   DOMAIN      466    469       POLY-GLY.
FT   DOMAIN      504    507       POLY-ALA.
FT   DOMAIN      512    518       POLY-ALA.
SQ   SEQUENCE   550 AA;  58930 MW;  F927C5E0 CRC32;


  Query Match             2.6%;  Score 7;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      47 LGGGGSG 53
           |||||||
Db     214 LGGGGSG 220


RESULT  13
CEA2_ECOLI
ID   CEA2_ECOLI     STANDARD;      PRT;   581 AA.
AC   P04419;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   COLICIN E2 (EC 3.1.21.1).
GN   COL OR CEAB.
OS   Escherichia coli.
OG   Plasmid ColE2-P9.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85239907.
RA   COLE S.T., SAINT-JOANIS B., PUGSLEY A.P.;
RT   "Molecular characterisation of the colicin E2 operon and
RT   identification of its products.";
RL   Mol. Gen. Genet. 198:465-472(1985).
RN   [2]
RP   SEQUENCE OF 377-581 FROM N.A.
RX   MEDLINE; 85062845.
RA   LAU P.C.K., ROWSOME R.W., ZUKER M., VISENTIN L.P.;
RT   "Comparative nucleotide sequences encoding the immunity proteins and
RT   the carboxyl-terminal peptides of colicins E2 and E3.";
RL   Nucleic Acids Res. 12:8733-8745(1984).
RN   [3]
RP   SEQUENCE OF 478-581 FROM N.A.
RX   MEDLINE; 85215569.
RA   MASAKI H., TOBA M., OHTA T.;
RT   "Structure and expression of the ColE2-P9 immunity gene.";
RL   Nucleic Acids Res. 13:1623-1635(1985).
CC   -!- FUNCTION: THIS PLASMID-CODED BACTERICIDAL PROTEIN IS AN
CC       ENDONUCLEASE ACTIVE ON BOTH SINGLE-AND DOUBLE-STRANDED DNA BUT
CC       WITH UNDEFINED SPECIFICITY.
CC   -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC       AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
```

```
CC   -!- SIMILARITY: BELONGS TO THE NUCLEASE FAMILY OF COLICINS AND
CC       PYOSINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29885; AAA23068.1; -.
DR   EMBL; X01163; CAA25609.1; -.
DR   EMBL; X02227; CAA26145.1; -.
DR   PIR; A00789; NDECE2.
KW   Plasmid; Bacteriocin; Colicin; Toxin; Hydrolase; Endonuclease.
FT   CONFLICT    429    431       QER -> RS (IN REF. 2).
FT   CONFLICT    473    473       A -> R (IN REF. 2).
FT   CONFLICT    504    504       F -> L (IN REF. 2).
SQ   SEQUENCE   581 AA;  61629 MW;  72E7E788 CRC32;


  Query Match             2.6%;  Score 7;  DB 1;  Length 581;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     49 GGGSGSG 55
          |||||||
Db     47 GGGSGSG 53


RESULT  14
CEA3_ECOLI
ID   CEA3_ECOLI     STANDARD;      PRT;   551 AA.
AC   P00646;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   COLICIN E3 (EC 3.1.-.-) (COLICIN E3 A CHAIN) (RIBONUCLEASE).
GN   CEAC.
OS   Escherichia coli.
OG   Plasmid ColE3-CA38.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85210906.
RA   MASAKI H., OHTA T.;
RT   "Colicin E3 and its immunity genes.";
RL   J. Mol. Biol. 182:217-227(1985).
RN   [2]
RP   SEQUENCE OF 370-551 FROM N.A.
RX   MEDLINE; 85062845.
RA   LAU P.C.K., ROWSOME R.W., ZUKER M., VISENTIN L.P.;
RT   "Comparative nucleotide sequences encoding the immunity proteins and
RT   the carboxyl-terminal peptides of colicins E2 and E3.";
RL   Nucleic Acids Res. 12:8733-8745(1984).
RN   [3]
RP   SEQUENCE OF 428-551 FROM N.A.
RX   MEDLINE; 83220760.
RA   MOCK M., MIYADA C.G., GUNSALUS R.P.;
RT   "Nucleotide sequence for the catalytic domain of colicin E3 and its
RT   immunity protein. Evidence for a third gene overlapping colicin.";
RL   Nucleic Acids Res. 11:3547-3557(1983).
RN   [4]
RP   SEQUENCE OF 445-551 FROM N.A.
RX   MEDLINE; 83105715.
RA   MASAKI H., OHTA T.;
RT   "A plasmid region encoding the active fragment and the inhibitor
RT   protein of colicin E3-CA38.";
RL   FEBS Lett. 149:129-132(1982).
RN   [5]
RP   SEQUENCE OF 455-551.
RX   MEDLINE; 79088682.
RA   SUZUKI K., IMAHORI K.;
RT   "Amino acid sequence of an active fragment (T2A) of colicin E3.";
RL   J. Biochem. 84:1031-1039(1978).
CC   -!- FUNCTION: INACTIVATES RIBOSOMES BY HYDROLYZING 16S RNA IN 30S
CC       RIBOSOMES AT A SPECIFIC SITE.
CC   -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC       AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC   -!- SIMILARITY: TO COLICIN E6 AND CLOACIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J01574; AAA88416.1; -.
DR   EMBL; X02397; CAA26241.1; -.
DR   EMBL; X01162; CAA25607.1; -.
DR   PIR; A00792; NRECE3.
DR   PIR; S07269; S07269.
KW   Plasmid; Bacteriocin; Colicin; Toxin; Hydrolase; Endonuclease.
FT   DOMAIN      455    551       RIBOSOME INACTIVATING ACTIVITY.
FT   DOMAIN      530    551       BINDING OF IMMUNITY PROTEIN.
SQ   SEQUENCE   551 AA;  57963 MW;  9DF18526 CRC32;


  Query Match             2.6%;  Score 7;  DB 1;  Length 551;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     49 GGGSGSG 55
          |||||||
Db     47 GGGSGSG 53


RESULT  15
CEA6_ECOLI
ID   CEA6_ECOLI     STANDARD;      PRT;   551 AA.
AC   P17999;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   COLICIN E6 (EC 3.1.-.-) (RIBONUCLEASE).
OS   Escherichia coli.
OG   Plasmid ColE6-CT14.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90078082.
RA   AKUTSU A., MASAKI H., OHTA T.;
RT   "Molecular structure and immunity specificity of colicin E6, an
RT   evolutionary intermediate between E-group colicins and cloacin
RT   DF13.";
RL   J. Bacteriol. 171:6430-6436(1989).
RN   [2]
RP   SEQUENCE OF 377-551 FROM N.A.
RX   MEDLINE; 89364708.
RA   LAU P.C.K., CONDIE J.A.;
RT   "Nucleotide sequences from the colicin E5, E6 and E9 operons:
RT   presence of a degenerate transposon-like structure in the ColE9-J
RT   plasmid.";
RL   Mol. Gen. Genet. 217:269-277(1989).
CC   -!- FUNCTION: INACTIVATES RIBOSOMES BY HYDROLYZING 16S RNA IN 30S
CC       RIBOSOMES AT A SPECIFIC SITE.
CC   -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC       AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC   -!- SIMILARITY: TO COLICIN E3 AND CLOACIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X15856; CAA33855.1; -.
DR EMBL; M31808; AAA23080.1; -.
KW Plasmid; Bacteriocin; Colicin; Toxin; Hydrolase; Endonuclease.
FT DOMAIN 455 551 RIBOSOME INACTIVATING ACTIVITY.
FT DOMAIN 530 551 BINDING OF IMMUNITY PROTEIN (BY
FT SIMILARITY).
SQ SEQUENCE 551 AA; 58011 MW; 34FB6FC9 CRC32;

Query Match 2.6%; Score 7; DB 1; Length 551;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
     49 GGGSGSG 55
        |||||||
Db   47 GGGSGSG 53
```

Search completed: February 11, 2000, 05:25:12
Job time: 516 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 05:40:21 ; Search time 41.27 Seconds
(without alignments)
451.923 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 269
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: OLIGO

Searched: 225878 seqs, 69334122 residues

Database : SPTREMBL_12:*

Word size : 0

Number of hits that pass the threshold : 225878
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	3.0	262	2	Q44188	Q44188 agrobacteri
2	8	3.0	529	2	P74332	P74332 synechocyst
3	8	3.0	1404	4	O15080	O15080 homo sapien
4	8	3.0	816	5	Q24538	Q24538 drosophila
5	8	3.0	598	5	Q24732	Q24732 drosophila
6	8	3.0	548	5	Q19496	Q19496 caenorhabdi
7	8	3.0	530	5	O76918	O76918 drosophila
8	8	3.0	255	8	Q9XKD9	Q9xkd9 dicyema mis
9	8	3.0	322	10	O65517	O65517 arabidopsis
10	8	3.0	2946	10	O64634	O64634 arabidopsis
11	8	3.0	207	10	Q43522	Q43522 lycopersico
12	8	3.0	256	10	O24568	O24568 zea mays (m
13	8	3.0	288	10	Q96463	Q96463 hordeum vul
14	8	3.0	330	12	Q89903	Q89903 mouse cytom
15	8	3.0	252	12	O36415	O36415 alcelaphine
16	7	2.6	434	1	O74038	O74038 cenarchaeum
17	7	2.6	317	1	O57894	O57894 pyrococcus
18	7	2.6	411	1	O58003	O58003 pyrococcus
19	7	2.6	679	1	O59078	O59078 pyrococcus
20	7	2.6	510	1	O59101	O59101 pyrococcus
21	7	2.6	852	1	Q9YDL5	Q9ydl5 aeropyrum p
22	7	2.6	301	1	Q9YDA6	Q9yda6 aeropyrum p
23	7	2.6	343	1	Q9YAA7	Q9yaa7 aeropyrum p
24	7	2.6	523	2	O54200	O54200 streptomyce
25	7	2.6	506	2	O31953	O31953 bacillus su
26	7	2.6	293	2	O51095	O51095 borrelia bu
27	7	2.6	458	2	O05591	O05591 mycobacteri
28	7	2.6	273	2	O83115	O83115 treponema p
29	7	2.6	280	2	O83475	O83475 treponema p
30	7	2.6	248	2	O84240	O84240 chlamydia t
31	7	2.6	489	2	O85724	O85724 streptomyce
32	7	2.6	448	2	O24842	O24842 acinetobact
33	7	2.6	242	2	Q46297	Q46297 clostridium
34	7	2.6	288	2	P73802	P73802 synechocyst
35	7	2.6	696	2	O54201	O54201 streptomyce
36	7	2.6	576	2	Q51604	Q51604 escherichia
37	7	2.6	173	2	Q9ZLL5	Q9zll5 helicobacte
38	7	2.6	308	2	Q51635	Q51635 shigella so
39	7	2.6	308	2	Q51636	Q51636 shigella so
40	7	2.6	308	2	Q51639	Q51639 escherichia
41	7	2.6	466	2	O87185	O87185 streptococc
42	7	2.6	499	2	Q9XB47	Q9xb47 escherichia
43	7	2.6	168	2	Q9X972	Q9x972 streptococc
44	7	2.6	640	2	Q9WZB2	Q9wzb2 thermotoga
45	7	2.6	889	3	O42723	O42723 emericella

ALIGNMENTS

RESULT 1
Q44188
ID Q44188 PRELIMINARY; PRT; 262 AA.
AC Q44188;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.
GN OATA.
OS Agrobacterium radiobacter.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T1305 LAC9;
RX MEDLINE; 96425889.
RA TIBURTIUS A., DE LUCA N., HUSSAIN H., JOHNSTON A.W.B.;
RT "Expression of the exoY gene, required for exopolysaccharide synthesis
RT in Agrobacterium, is activated by the regulatory ros gene.";
RL Microbiology 142:2621-2629(1996).
DR EMBL; X95394; CAA64678.1; -.
DR HSSP; P16932; 1DGE.
DR PFAM; PF00202; aminotran_3; 1.
KW Transferase.
SQ SEQUENCE 262 AA; 28124 MW; DF7A4568 CRC32;

Query Match 3.0%; Score 8; DB 2; Length 262;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LKGSPNVI 11
||||||||
Db 244 LKGSPNVI 251

RESULT 2
P74332
ID P74332 PRELIMINARY; PRT; 529 AA.
AC P74332;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 58.0 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.

```
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90914; BAA18426.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   529 AA;  57992 MW;  15123FB3 CRC32;
```

[redacted]

```
  Query Match             3.0%;  Score 8;  DB 2;  Length 529;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     43 SLPLLGGG 50
          ||||||||
Db    260 SLPLLGGG 267
```

```
RESULT   3
O15080
ID   O15080      PRELIMINARY;      PRT;  1404 AA.
AC   O15080;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   KIAA0375.
GN   KIAA0375.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97349984.
RA   NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA   TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
DR   EMBL; AB002373; BAA20830.1; -.
SQ   SEQUENCE   1404 AA;  149264 MW;  675DE431 CRC32;
```

```
  Query Match             3.0%;  Score 8;  DB 4;  Length 1404;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     37 LALLPFSL 44
          ||||||||
Db   1163 LALLPFSL 1170
```

```
RESULT   4
Q24538
ID   Q24538      PRELIMINARY;      PRT;  816 AA.
AC   Q24538;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   STEROID RECEPTOR BETA FTZ-F1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91227912.
RA   LAVORGNA G., UEDA H., CLOS J., WU C.;
RT   "FTZ-F1, a steroid hormone receptor-like protein implicated in the
RT   activation of fushi tarazu.";
RL   Science 252:848-851(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93219413.
RA   LAVORGNA G., KARIM F.D., THUMMEL C.S., WU C.;
RT   "Potential role for a FTZ-F1 steroid receptor superfamily member in
RT   the control of Drosophila metamorphosis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:3004-3008(1993).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; M98397; AAA28915.1; -.
DR   HSSP; P19793; 2NLL.
DR   FLYBASE; FBgn0001078; ftz-f1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR   PFAM; PF00104; hormone_rec; 1.
DR   PFAM; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
KW   Hormone; Receptor; Transcription regulation; DNA-binding;
KW   Nuclear protein; Zinc-finger.
SQ   SEQUENCE   816 AA;  87415 MW;  949464D9 CRC32;
```

```
  Query Match             3.0%;  Score 8;  DB 5;  Length 816;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 GGGGSGSG 55
          ||||||||
Db     96 GGGGSGSG 103
```

```
RESULT   5
Q24732
ID   Q24732      PRELIMINARY;      PRT;  598 AA.
AC   Q24732;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLASS.
GN   GLASS.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95327686.
RA   O'NEILL E.M., ELLIS M.C., RUBIN G.M., TJIAN R.;
RT   "Functional domain analysis of glass, a zinc-finger-containing
RT   transcription factor in Drosophila.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:6557-6561(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   LIU H., MA C., MOSES K.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U39746; AAA85272.1; -.
DR   HSSP; P08046; 1A1F.
DR   TRANSFAC; T02282; -.
DR   FLYBASE; FBgn0015211; Dvir\gl.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 5.
DR   PFAM; PF00096; zf-C2H2; 5.
DR   PRINTS; PR00048; ZINCFINGER.
KW   Zinc-finger; Metal-binding; DNA-binding.
SQ   SEQUENCE   598 AA;  61773 MW;  87921C20 CRC32;
```

```
  Query Match             3.0%;  Score 8;  DB 5;  Length 598;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db     244 GGGGSGSG 251


RESULT   6
Q19496
ID   Q19496      PRELIMINARY;      PRT;   548 AA.
AC   Q19496;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 63.5 KD PROTEIN.
     F16H11.5.
     Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WU X.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
     STRAIN=BRISTOL N2;
     WATERSTON R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U55376; AAA98005.1; -.
DR   HSSP; P03372; 1HCP.
DR   PFAM; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
KW   Hypothetical protein.
FT   DOMAIN      139    144       POLY-GLN.
SQ   SEQUENCE   548 AA;  63545 MW;  DD7D3BA5 CRC32;


  Query Match             3.0%;  Score 8;  DB 5;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     153 KDSSGNET 160
           ||||||||
Db     109 KDSSGNET 116


RESULT   7
O76918
ID   O76918      PRELIMINARY;      PRT;   530 AA.
AC   O76918;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   EG:96G10.7 PROTEIN.
GN   EG:96G10.7.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BOLSHAKOV V., BORKOVA D., MINANA B., KAFATOS F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   BENOS P.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL024484; CAA19668.1; -.
SQ   SEQUENCE   530 AA;  54644 MW;  071B0D78 CRC32;


  Query Match             3.0%;  Score 8;  DB 5;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db      65 GGGGSGSG 72


RESULT   8
Q9XKD9
ID   Q9XKD9      PRELIMINARY;      PRT;   255 AA.
AC   Q9XKD9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME C OXIDASE SUBUNIT III.
GN   COIII.
OS   Dicyema misakiense.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Mesozoa; Rhombozoa; Dicyemida; Dicyemidae;
OC   Dicyema.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99150413.
RA   WATANABE K.I., BESSHO Y., KAWASAKI M., HORI H.;
RT   "Mitochondrial genes are found on minicircle DNA molecules in the
RT   mesozoan animal Dicyema.";
RL   J. Mol. Biol. 286:345-650(1999).
DR   EMBL; AB011834; BAA76312.1; -.
KW   Mitochondrion.
SQ   SEQUENCE   255 AA;  28779 MW;  865DC7A1 CRC32;


  Query Match             3.0%;  Score 8;  DB 8;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 PFSLPLLG 48
           ||||||||
Db     127 PFSLPLLG 134


RESULT   9
O65517
ID   O65517      PRELIMINARY;      PRT;   322 AA.
AC   O65517;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 34.1 KD PROTEIN.
GN   F23E13.150.
```

```
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., HILBERT H., BRAUN M., HOLZER E., BRANDT A., DUESTERHOEFT A.,
RA   HOHEISEL J., JESSE T., HEIJNEN L., VOS P., MEWES H.W., MAYER K.,
RA   SCHUELLER C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL022141; CAA18132.1; -.
DR   MENDEL; 29098; Arath;2109;29098.
KW   Hypothetical protein.
SQ   SEQUENCE   322 AA;  34144 MW;  D8212ADB CRC32;

  Query Match             3.0%;  Score 8;  DB 10;  Length 322;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 GGGGSGSG 55
          ||||||||
Db    149 GGGGSGSG 156


RESULT  10
O64634
ID   O64634      PRELIMINARY;      PRT;  2946 AA.
AC   O64634;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CDC4 LIKE PROTEIN.
GN   F17K2.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA   BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC003680; AAC06163.1; -.
DR   HSSP; P04002; 1WFA.
DR   PFAM; PF00400; WD40; 1.
SQ   SEQUENCE   2946 AA;  321930 MW;  E6094427 CRC32;

  Query Match             3.0%;  Score 8;  DB 10;  Length 2946;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     47 LGGGGSGS 54
          ||||||||
Db   1507 LGGGGSGS 1514


RESULT  11
Q43522
ID   Q43522      PRELIMINARY;      PRT;  207 AA.
AC   Q43522;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   TFM5 GENE.
GN   TFM5.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC   Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. UC82B; TISSUE=FRUIT;
RX   MEDLINE; 97201476.
RA   SANTINO C.G., STANFORD G.L., CONNER T.W.;
RT   "Developmental and transgenic analysis of two tomato fruit enhanced
RT   genes.";
RL   Plant Mol. Biol. 33:405-416(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. UC82B; TISSUE=FRUIT;
RA   CONNOR T.W.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X95262; CAA64559.1; -.
DR   HSSP; P24337; 1HYP.
DR   MENDEL; 16224; Lyces;1531;16224.
DR   PFAM; PF00279; LTP; 1.
SQ   SEQUENCE   207 AA;  18265 MW;  B102F0EB CRC32;

  Query Match             3.0%;  Score 8;  DB 10;  Length 207;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 GGGGSGSG 55
          ||||||||
Db     45 GGGGSGSG 52


RESULT  12
O24568
ID   O24568      PRELIMINARY;      PRT;  256 AA.
AC   O24568;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCINE-RICH PROTEIN PRECURSOR.
GN   GRP3.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Zea.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DK105; TISSUE=ROOT;
RX   MEDLINE; 98187261.
RA   GODDEMEIER M.L., WULFF D., FEIX G.;
RT   "Root-specific expression of a Zea mays gene encoding a novel glycine-
RT   rich protein, zmGRP3.";
RL   Plant Mol. Biol. 36:799-802(1998).
DR   EMBL; Y07781; CAA69104.1; -.
DR   MENDEL; 27037; Zeama;343;27037.
KW   Signal.
FT   SIGNAL       1     23       POTENTIAL.
SQ   SEQUENCE   256 AA;  21712 MW;  0BF5ED59 CRC32;

  Query Match             3.0%;  Score 8;  DB 10;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 GGGGSGSG 55
          ||||||||
Db    242 GGGGSGSG 249
```

```
RESULT  13
Q96463
ID   Q96463      PRELIMINARY;      PRT;   288 AA.
AC   Q96463;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MYB4 TRANSCRIPTION FACTOR (FRAGMENT).
GN   MYB4.
OS   Hordeum vulgare (Barley).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Hordeum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. HIMALAYA; TISSUE=ENDOSPERM;
RA   GUBLER F., ROBERTS J.K., JACOBSEN J.;
RL   Plant Physiol. 113:306-306(1997).
DR   EMBL; X99973; CAA68235.1; -.
DR   HSSP; P01103; 1POM.
DR   MENDEL; 8597; Horvu;Myb;8597.
DR   PFAM; PF00249; myb_DNA-binding; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   288 AA;  31044 MW;  FB45FD67 CRC32;


  Query Match             3.0%;  Score 8;  DB 10;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db     265 GGGGSGSG 272


RESULT  14
Q89903
ID   Q89903      PRELIMINARY;      PRT;   330 AA.
AC   Q89903;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   E1 PROTEIN.
GN   E1.
OS   Mouse cytomegalovirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Muromegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90219186.
RA   BUEHLER B., KEIL G.M., WEILAND F., KOSZINOWSKI U.H.;
RT   "Characterization of the murine cytomegalovirus early transcription
RT   unit e1 that is induced by immediate-early proteins.";
RL   J. Virol. 64:1907-1919(1990).
DR   EMBL; L07320; AAA45907.1; -.
DR   EMBL; M35146; AAA45908.1; -.
SQ   SEQUENCE   330 AA;  34605 MW;  E3FC3225 CRC32;


  Query Match             3.0%;  Score 8;  DB 12;  Length 330;
  Best Local Similarity   100.0%;  Pred. No. 8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db     238 GGGGSGSG 245


RESULT  15
O36415
ID   O36415      PRELIMINARY;      PRT;   252 AA.
AC   O36415;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CAPSID PROTEIN.
OS   Alcelaphine herpesvirus 1 (wildebeest herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C500;
RX   MEDLINE; 97404659.
RA   ENSSER A., PFLANZ R., FLECKENSTEIN B.;
RT   "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL   J. Virol. 71:6517-6525(1997).
DR   EMBL; AF005370; AAC58112.1; -.
SQ   SEQUENCE   252 AA;  24554 MW;  6B323798 CRC32;


  Query Match             3.0%;  Score 8;  DB 12;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      48 GGGGSGSG 55
           ||||||||
Db     96 GGGGSGSG 103


Search completed: February 11, 2000, 05:58:00
Job time: 1059 sec
```

```
OM of: US-09-060-609-2 to: Issued_Patents_NA:*   out_format : pfs

Date: Feb 11, 2000  6:00 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09060609/runat_09022000_150438_25190/app_query.fasta.1
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09060609 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-060-609-2
Query length: 269
Database: Issued_Patents_NA:*
Database sequences: 207703
Database length: 57918364
Search time (sec): 41.230000



score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-295-814E-1 +   88.50  140.79  2.03      2028  !
/cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:PCT-US93-01959-1 +   88.50  140.79  2.03      2028
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-700-013B-7 +   87.00  140.68  2.06      1597  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-700-013B-20 +   87.00  135.73  3.88      2397  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-700-013B-18 +   86.00  133.72  5.02      2397  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-3 +   85.00  139.76  2.31      1238  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-5 +   85.00  142.19  1.69      1014  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-177-109A-1 +   84.50  130.28  7.81      2483  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-687-706-1 +   84.50  130.28  7.81      2483  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-121-057-2 +   84.00  123.23  19.29     4079  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-509-187D-2 +   84.00  123.23  19.29     4079  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-09-121-396-2 +   84.00  123.23  19.29     4079  !
/cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:PCT-US93-09704A-2 +   84.00  123.23  19.29     407
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-276-967-1 +   83.00  113.34  68.52     7785  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-828-832-2 +   82.00  141.59  1.83      649  !  S
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-2 -   82.00   97.27  538.56    24701  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-3 +   82.00   97.27  538.56    24701  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-6 -   82.00  109.61  110.60    8967  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-9 +   82.00  109.61  110.60    8967  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-26 -   82.00  119.84  29.76     3870  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-60 -   82.00   97.27  538.56    24701
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-61 +   82.00   97.27  538.56    24701
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-64 -   82.00  109.61  110.60    8967  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-853-659A-67 +   82.00  109.61  110.60    8967  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-665-617-1 +   81.50  136.02  3.74      944  !  S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-631-200-10 -   81.00  136.65  3.45      826  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-829-553-10 -   81.00  136.65  3.45      826  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-922-267A-10 -   81.00  136.65  3.45      826  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-936-707A-10 -   81.00  136.65  3.45      826  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-936-706A-10 -   81.00  136.65  3.45      826  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-796-414B-6 -   81.00  122.73  20.54     2588  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-804-227C-1 -   81.00   88.42  1.7e+03   43280  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-717-123-1 +   81.00  134.99  4.26      946  !  S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-668-648-5 -   80.50  126.38  12.88     1767  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-241-853-1 +   80.50  116.52  45.58     3969  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-614-770A-1 +   80.50   85.84  2.3e+03   49272  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-850-917-1 +   80.50  116.52  45.58     3969  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-1 +   80.50  133.60  5.10      977  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-6 +   80.50  134.43  4.58      912  !
/cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-429-998-5 -   80.50  126.38  12.88     1767  !
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-431-333-5 -   80.50  126.38  12.88     1767  !
/cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:PCT-US91-02321-5 -   80.50  126.38  12.88     1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-463-931-1 +   80.00  123.03  19.77     2141
/cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-666-392A-1 +   80.00  127.92  10.56     1433
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-199-780-2 +   79.00  112.49  76.44     4314

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-295-814E-1

seq_documentation_block:
; Sequence 1, Application US/08295814E
; Patent No. 5658786
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Kelli E.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  Hartig, Paul R.
;    APPLICANT:  Weinshank, Richard L.
;    TITLE OF INVENTION:  DNA ENCODING TAURINE AND GABA
;    TITLE OF INVENTION:  TRANSPORTERS AND USES THEREOF
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/295,814E
;      FILING DATE:  DECEMBER 19, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40558-B-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2028 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FRAGMENT TYPE:  N-terminal
;    IMMEDIATE SOURCE:
;      LIBRARY:  rat brain
;      CLONE:  rB14b
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  126..1932
;      OTHER INFORMATION:
US-08-295-814E-1


alignment_scores:
            Quality:   88.50              Length:    248
              Ratio:  0.763                Gaps:     15
 Percent Similarity:  46.774   Percent Identity: 21.774

alignment_block:
 US-09-060-609-2 x US-08-295-814E-1   ..

 Align seg 1/1  to: US-08-295-814E-1  from: 1  to: 2028

      52 SerGlySerGlyGluLysValSerValSerLysMetAla...AlaAlaTr 67
         :::|||:::::::::      :::   |||      |||   |||:::||
```

```
   1057 ACAGGGACTGCGTCGCCCTTTGCATTCTCAACAGCAGCACCAGCTTCGTG 1106

     67 pProSerGlyProSerAlaProGluAlaValThrAla..... 79
        ||||:::   ||||||:::|||:::   :::::::::
   1107 GCCGGGTTTGCCATCTTCTCCATCCTGGGCTTCATGTCTCAGGAGCAGGG 1156

     80 .............ArgLeuVal......GlyValLeuTrpPheValSer 89
                    |||||||:::      ::::::   |||      |||
   1157 CGTACCCATATCTGAGGTTGCTGAATCAGGCCCTGGCCTGGCATTCATCG 1206

     90 ValThrThrGlyProTrp.......................GlyAlaVa 98
        |||         |||                         |||   ::
   1207 CCTACCCTCGAGCTGTGGTGATGTTACCTTTCTCGCCTTTGTGGGCCTGC 1256

     98 lAlaThrSerAla.GlyGlyGluGluSerLeuLys.........CysGlu 111
        :::::::|||::: |||||||:::   :::   :::          |||
   1257 TGTTTCTTCTTCATGGTGGTTCTCCTGGGACTAGACAGCCAGTTTGT... 1303

    112 AspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAl 128
                          :::|||:::   |||:::
   1304 .....................GTGTGTAGAAAGCCTCGT........... 1321

    128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                            :::      ::::::::::
   1322 .....................GACAGCGCTGGTGGACATGTATCC.... 1345

    145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
           |||::::::     ::::::   :::|||   :::::::|||
   1346 ........CCGGGTGTTCCGTAAGAAGAACCGGAGGGA...GATTCTCAT 1384

    162 .PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnV 178
         :::      :::   |||:::||||||      :::         :::
   1385 CCTCATCGTGTCTGTCGTCTCTTTCTTCATCGGGCTCATTATGCTCACAG 1434

    178 alAsnGly...TyrSerTyrLys.......................... 184
        :::|||   |||   ::::::
   1435 AGGGCGGCATGTACGTGTTCCAGCTCTTCGACTACTATGCGGCCAGTGGC 1484

    185 ..............ValAlaValAlaLeuSerLeuPheLeuGlyTrpLe 196
                      ||||||:::      ||||||   :::::::|||::
   1485 ATGTGTCTTCTCTTTGTGGCCATCTTTGAGTCCCTCTGTGTGGCTTGGGT 1534

    196 u...GlyAlaAspArgPheTyr.................LeuGlyTyrP 206
        :   |||||||:::|||||||||                  :::|||||||
   1535 TTACGGAGCCAGCCGCTTCTATGACAACATTGAAGATATGATTGGGTACA 1584

    206 roAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGly 222
                 |||:::|||:::|||   :::   |||
   1585 AGCCGTGGCCTCTTATCAAATACTGTTGGCTCTTTTTCACGCCAGCTGTG 1634

    223 SerLeuIleAspPheIleLeuIleSerMetGlnIleValGlyProSerAs 239
           |||      |||::::::      ::::::   :::      :::
   1635 TGCCTGGCAACCTTCCTGTTCTCCCTGATCAAATACACGCCACTGACCTA 1684

    239 pGlySerSerTyrIleIleAspTyrTyrGlyThrArgLeu 252
         ::::::::::|||         ::::::|||      |||
   1685 CAACAAGAAGTACACATATCCATGGTGGGGGGATGCCCTG 1724
```

```
seq_name: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:PCT-US93-01959-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9301959
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, E. Kelli
;    APPLICANT:  Borden, A. Laurence
;    APPLICANT:  Hartig, R. Paul
;    APPLICANT:  Weinshank, L. Richard
;    TITLE OF INVENTION:  DNA ENCODING TAURINE AND GABA
;    TITLE OF INVENTION:  TRANSPORTERS AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/01959
;      FILING DATE:  19930304
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40558A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-977-9550
;      TELEFAX:  212-664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2028 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FRAGMENT TYPE:  N-terminal
;    IMMEDIATE SOURCE:
;      LIBRARY:  rat brain
;      CLONE:  rB14b
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  126..1932
;      OTHER INFORMATION:
PCT-US93-01959-1


alignment_scores:
            Quality:   88.50              Length:    248
              Ratio:  0.763                Gaps:     15
 Percent Similarity:  46.774   Percent Identity: 21.774

alignment_block:
 US-09-060-609-2 x PCT-US93-01959-1   ..

 Align seg 1/1  to: PCT-US93-01959-1  from: 1  to: 2028
```

```
      52 SerGlySerGlyGluLysValSerValSerLysMetAla...AlaAlaTr 67
         :::|||:::::::::      :::   |||      |||   |||:::||
   1057 ACAGGGACTGCGTCGCCCTTTGCATTCTCAACAGCAGCACCAGCTTCGTG 1106

     67 pProSerGlyProSerAlaProGluAlaValThrAla..... 79
        ||||:::   ||||||:::|||:::   :::::::::
   1107 GCCGGGTTTGCCATCTTCTCCATCCTGGGCTTCATGTCTCAGGAGCAGGG 1156

     80 .............ArgLeuVal......GlyValLeuTrpPheValSer 89
                    ||||||:::      ::::::   |||      |||
   1157 CGTACCCATATCTGAGGTTGCTGAATCAGGCCCTGGCCTGGCATTCATCG 1206

     90 ValThrThrGlyProTrp.......................GlyAlaVa 98
        |||         |||                         |||   ::
   1207 CCTACCCTCGAGCTGTGGTGATGTTACCTTTCTCGCCTTTGTGGGCCTGC 1256

     98 lAlaThrSerAla.GlyGlyGluGluSerLeuLys.........CysGlu 111
        :::::::|||::: |||||||:::   :::   :::          |||
   1257 TGTTTCTTCTTCATGGTGGTTCTCCTGGGACTAGACAGCCAGTTTGT... 1303

    112 AspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAl 128
                          :::|||:::   |||:::
```

```
    1304 .....GTGTGTAGAAAGCCTCGT..... 1321

     128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                               :::      :::||||||||||
    1322 .....GACAGCGCTGGTGGACATGTATCC.... 1345

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
              |||::::::      ::::::   :::|||   :::::::|||
    1346 ........CCGGGTGTTCCGTAAGAAGAACCGGAGGGA...GATTCTCAT 1384

     162 .PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnV 178
          :::      :::   |||:::||||||      :::         :::
    1385 CCTCATCGTGTCTGTCGTCTCTTTCTTCATCGGGCTCATTATGCTCACAG 1434

     178 alAsnGly...TyrSerTyrLys.......................... 184
           :::|||   |||   ::::::
    1435 AGGGCGGCATGTACGTGTTCCAGCTCTTCGACTACTATGCGGCCAGTGGC 1484

     185 ..............ValAlaValAlaLeuSerLeuPheLeuGlyTrpLe 196
                      ||||||:::      ||||||   :::::::|||::
    1485 ATGTGTCTTCTCTTTGTGGCCATCTTTGAGTCCCTCTGTGTGGCTTGGGT 1534

     196 u...GlyAlaAspArgPheTyr.....:..........LeuGlyTyrP 206
          :   ||||||:::|||||||||                ::::||||||
    1535 TTACGGAGCCAGCCGCTTCTATGACAACATTGAAGATATGATTGGGTACA 1584

     206 roAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGly 222
               |||:::|||:::|||   :::   |||
    1585 AGCCGTGGCCTCTTATCAAATACTGTTGGCTCTTTTTCACGCCAGCTGTG 1634

     223 SerLeuIleAspPheIleLeuIleSerMetGlnIleValGlyProSerAs 239
           |||      |||::::::      ::::::   :::      :::
    1635 TGCCTGGCAACCTTCCTGTTCTCCCTGATCAAATACACGCCACTGACCTA 1684

     239 pGlySerSerTyrIleIleAspTyrTyrGlyThrArgLeu 252
          ::::::::::|||        ::::::|||      |||
    1685 CAACAAGAAGTACACATATCCATGGTGGGGGGATGCCCTG 1724

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-700-013B-7

seq_documentation_block:
; Sequence 7, Application US/08700013B
; Patent No. 5919653
;  GENERAL INFORMATION:
;    APPLICANT:  Albert, Vivian R.
;    APPLICANT:  Kowalski, Leslie R.Z.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  McKelvy, Jeffrey F.
;    TITLE OF INVENTION:  Human Glycine Transporter
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/700,013B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1597 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-700-013B-7


alignment_scores:
            Quality:   87.00             Length:    302
              Ratio:  0.617               Gaps:     17
 Percent Similarity:  46.689   Percent Identity: 22.517

alignment_block:
 US-09-060-609-2 x US-08-700-013B-7   ..

 Align seg 1/1  to: US-08-700-013B-7  from: 1  to: 1597

      25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
         ||||||||||:::|||:::::::|||      ::::::   |||
     553 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 602

      39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly.............. 51
          |||       |||||||||   |||          |||
     603 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 652

      52 .................SerGlySerGly.Glu................... 56
                         :::|||||||||
     653 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 702

      57 .LysValSerValSerLysMetAlaAlaAlaTrpPro.............. 68
         :::|||||||:::::::::::::|||:::      |||
     703 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 752

      69 .................SerGlyProSerAlaProGluAlaValThrAla 79
                         :::|||   :::            :::::::::
     753 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCAACTCT 802

      80 ArgLeuValGlyValLeuTrp............................. 86
           |||:::::::::   :::
     803 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 852

      87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                     ||||||||||||         ||||||||||:::
     853 CCTGTTTGCCTCCTTTGTGTCTGTACTA......CCCTGGGGCTCCTGCA 896

      99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp......... 112
           :::      :::      |||        |||:::|||
     897 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 937

     113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
         |||   :::   :::   :::   :::|||   |||||||||
     938 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA........ 979

     128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                       :::      |||   |||      ::::::::::
     980 ...............CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1013

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
         ::|||      :::::::::         |||:::::::|||:::|||
    1014 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1063

     162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
           :::|||:::|||      :::|||||||   :::      :::
    1064 GTCAGTGGAAGTGAG...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1110

     178 lAsnGlyTyrSerTyr............LysValAlaValAlaLeuSerL 191
             |||   :::|||            :::      :::|||||||   |
```

```
   1111 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1160

    191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
        ||||||||:::|||:::                :::   |||   :::
   1161 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1192

    208 Leu.....................GlyLeuLeuLysPheCysThrValGl 217
        |||                    |||   :::   :::   |||:::
   1193 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1242

    217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
         |||      :::      |||:::   :::::::::   ::::::   :
   1243 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1289

    234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
        ::   |||   :::   |||      |||:::|||
   1290 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1339

    251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
        :::|||||||      ::::::   :::::::::         ||||||::
   1340 AAACTCACGGATGCCACGGTGTGGAAAGATGCT......GCCACTCAGAT 1383

    267 uTyr 268
        ::::
   1384 TTTC 1387

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-700-013B-20

seq_documentation_block:
; Sequence 20, Application US/08700013B
; Patent No. 5919653
;  GENERAL INFORMATION:
;    APPLICANT:  Albert, Vivian R.
;    APPLICANT:  Kowalski, Leslie R.Z.
;    APPLICANT:  Borden, Laurence A.
;    APPLICANT:  McKelvy, Jeffrey F.
;    TITLE OF INVENTION:  Human Glycine Transporter
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/700,013B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2397 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-700-013B-20

alignment_scores:
            Quality:   87.00              Length:    302
              Ratio:  0.617                Gaps:     17
 Percent Similarity:  46.689   Percent Identity: 22.517

alignment_block:
 US-09-060-609-2 x US-08-700-013B-20   ..

 Align seg 1/1  to: US-08-700-013B-20  from: 1  to: 2397

      25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
         |||||||||:::|||:::::|||      ::::::   |||
     591 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 640

      39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly.............. 51
          |||      |||||||||   |||         |||
     641 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 690

      52 .................SerGlySerGly.Glu................... 56
                          ~       :::|||||||||
     691 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 740

      57 .LysValSerValSerLysMetAlaAlaAlaTrpPro.............. 68
          :::||||||:::::::::::::|||:::      |||
     741 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 790

      69 .................SerGlyProSerAlaProGluAlaValThrAla 79
                          :::|||   :::             :::::::::
     791 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCAACTCT 840

      80 ArgLeuValGlyValLeuTrp............................. 86
          |||::::::::::   :::
     841 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 890

      87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                     ||||||||||||         |||||||||:::
     891 CCTGTTTGCCTCCTTTGTGTCTGTACTA......CCCTGGGGCTCCTGCA 934

      99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp......... 112
           :::       :::       |||         |||:::|||
     935 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 975

     113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
          |||   :::   :::   :::   :::|||   |||||||||
     976 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA........ 1017

     128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                       :::       |||   |||       ::::::::::
    1018 ................CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1051

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
         ::|||       :::::::::          |||:::::::|||:::|||
    1052 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1101

     162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
          :::|||:::|||      :::||||||   :::      :::
    1102 GTCAGTGGAAGTGAG...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1148

     178 lAsnGlyTyrSerTyr............LysValAlaValAlaLeuSerL 191
          |||   :::|||            :::      :::||||||   |
    1149 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1198

     191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
         ||||||||:::|||:::                :::   |||   :::
    1199 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1230

     208 Leu.....................GlyLeuLeuLysPheCysThrValGl 217
         |||                    |||   :::   :::   |||:::
    1231 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1280

     217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
          |||      :::      |||:::   :::::::::   ::::::   :
```

```
    1281 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1327

     234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
         ::   |||   :::   |||      |||:::|||
    1328 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1377

     251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
         :::||||||      ::::::   :::::::::         |||||||::
    1378 AAACTCACGGATGCCACGGTGTGGAAAGATGCT......GCCACTCAGAT 1421

     267 uTyr 268
         ::::
    1422 TTTC 1425

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-700-013B-18

seq_documentation_block:
; Sequence 18, Application US/08700013B
; Patent No. 5919653
;   GENERAL INFORMATION:
;     APPLICANT:  Albert, Vivian R.
;     APPLICANT:  Kowalski, Leslie R.Z.
;     APPLICANT:  Borden, Laurence A.
;     APPLICANT:  McKelvy, Jeffrey F.
;     TITLE OF INVENTION:  Human Glycine Transporter
;     NUMBER OF SEQUENCES:  41
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dechert Price & Rhoads
;       STREET:  997 Lenox Drive, Building 3, Suite 210
;       CITY:  Lawrenceville
;       STATE:  NJ
;       COUNTRY:  USA
;       ZIP:  08543
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/700,013B
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bloom, Allen
;       REGISTRATION NUMBER:  29,135
;       REFERENCE/DOCKET NUMBER:  317743-108
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  609-520-3214
;       TELEFAX:  609-520-3259
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2397 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-700-013B-18


alignment_scores:
            Quality:   86.00              Length:    302
              Ratio:  0.614                Gaps:     17
 Percent Similarity:  46.358   Percent Identity: 22.517

alignment_block:
 US-09-060-609-2 x US-08-700-013B-18   ..

 Align seg 1/1  to: US-08-700-013B-18  from: 1  to: 2397

      25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla......Le 39
         |||||||||:::|||::::::|||      ::::::   |||
     591 ACTGGACTTCATCCTGTCCATGGTGGGGTACGCAGTGGGGCTGGGCAATG 640

      39 uLeuProPheSerLeuProLeuLeuGlyGlyGlyGly..... 51
         |||      |||||||||   |||         |||
     641 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGGAGGTGCTTTCCTC 690

      52 .................SerGlySerGly.Glu..... 56
                         :::||||||||||
     691 ATCCCTTACCTGATGATGCTGGCTCTGGCTGGATTACCCATCTTCTTCTT 740

      57 .LysValSerValSerLysMetAlaAlaAlaTrpPro..... 68
         :::||||||:::::::::::::|||:::      |||
     741 GGAGGTGTCGCTGGGCCAGTTTGCCAGCCAGGGACCAGTGTCTGTGTGGA 790

      69 .................SerGlyProSerAlaProGluAlaValThrAla 79
                         :::|||   :::           :::   :::
     791 AGGCCATCCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCATCTCT 840

      80 ArgLeuValGlyValLeuTrp..... 86
         |||:::::::::   :::
     841 GTCCTAATAGCCATATACTACAATGTGATTATTTGCTATACACTTTTCTA 890

      87 .............PheValSerValThrThrGlyProTrpGlyAlaValA 99
                     ||||||||||||         |||||||||:::
     891 CCTGTTTGCCTCCTTTGTGTCTGTACTA.....CCCTGGGGCTCCTGCA 934

      99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp..... 112
         :::      :::      |||         |||:::|||
     935 ACAACCCTTGGAATACGCCAGAA.........TGCAAAGATAAAACCAAA 975

     113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
         |||   :::   :::   :::   :::|||   |||||||||
     976 CTTTTATTAGATTCCTGTGTTATCAGTGACCATCCCAAAATA......... 1017

     128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
                     :::      |||   |||      ::::::::::
    1018 ...............CAGATCAAGAACTCGACTTTCTGCATGACCGCTT 1051

     145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
         ::|||    :::::::::          |||::::::|||:::|||
    1052 ATCCCAACGTGACAATGGTTAATTTCACCAGCCAGGCCAATAAGACATTT 1101

     162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
            :::|||:::|||      :::||||||   :::      :::
    1102 GTCAGTGGAAGTGAA...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1148

     178 lAsnGlyTyrSerTyr.....LysValAlaValAlaLeuSerL 191
            |||   :::|||           :::      :::||||||   |
    1149 TGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1198

     191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
         ||||||||:::|||:::                  :::   |||   :::
    1199 TCTTCCTGGCTTGGGTC.................ATTGTGTATGCATCG 1230

     208 Leu.....GlyLeuLeuLysPheCysThrValGl 217
         |||          |||   :::   :::   |||:::
    1231 TTGGCTAAAGGAATCAAGACTTCAGGAAAAGTGGTGTACTTCACGGCCAC 1280

     217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
         |||      :::      |||:::   :::::::::   ::::::   :
    1281 GTTCCCGTATGTC...GTACTCGTGATCCTCCTCATCCGAGGAGTCACCC 1327

     234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
         ::   |||   :::   |||      |||:::|||
    1328 TGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAGTGGGAG 1377

     251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
         :::||||||      ::::::   :::::::::         |||||||::
    1378 AAACTCACGGATGCCACGGTGTGGAAAGATGCT......GCCACTCAGAT 1421

     267 uTyr 268
         ::::
```

```
   1422 TTTC 1425

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-3

seq_documentation_block:
; Sequence 3, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  56...1069
;      OTHER INFORMATION:
US-08-414-657D-3


alignment_scores:
           Quality:   85.00              Length:    130
             Ratio:  1.308                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 27.692

alignment_block:
 US-09-060-609-2 x US-08-414-657D-3   ..

 Align seg 1/1  to: US-08-414-657D-3  from: 1  to: 1238

     101 SerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLysValGly.. 116
         :::::::|||   :::   ||||||||||||||         |||
     743 ACCACAGGACGACAAGCTTCCCTCAAATGTGAAGCCTCAGCGGTGCCTGC 792

     117 .........GlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130
                 ::::::        |||   :::|||||:::|||:::
     793 ACCTGACTTTGAGTGGTACCGGGATGACACCAGGATAAACAGTGCAAACG 842

     130 lnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPhePro 146
               ::::::   |||:::                |||   :::
     843 GCCTTGAGATTAAGAGCACTGAG.........GGCCAGTCCTCCCTGACG 883

     147 AlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheTh 163
         :::   |||:::|||   ::::::      ||||||   |||      ::
     884 GTGACCAACGTCACT...GAGGAACACTACGGCAACTATACCTGTGTGGC 930

     163 rGlyAsnGluValGly.....................PhePheLysProI 173
         ::::|||::::::|||                     :::|||:::|||
     931 TGCCAACAAGCTCGGCGTCACCAATGCCAGCCTAGTCCTTTTCAGACCCG 980

     173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
           |||   |||::::::||||||   |||   ::::::||||||   |||
     981 GGTCGGTGAGAGGAATCAACGGA...TCCATCAGTCTGGCCGTACCACTG 1027

     190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
                       ||||||
    1028 ..............TGGCTG............................. 1033

     206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCys 219
           |||:::        ||||||   ::::::   |||
    1034 ....CTGGCAGCGTCCCTGTTCTGCCTTCTCAGCAAATGT 1069

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-5

seq_documentation_block:
; Sequence 5, Application US/08414657D
; Patent No. 5861283
;  GENERAL INFORMATION:
;    APPLICANT:  Levitt, Pat
;    APPLICANT:  Pimenta, Aurea
;    APPLICANT:  Fischer, Itzhak
;    APPLICANT:  Zhukareva, Victoria
;    TITLE OF INVENTION:  Limbic System-Associated Membrane
;    TITLE OF INVENTION:  Protein and DNA
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/414,657D
;      FILING DATE:  31-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  317743-102
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1014 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...1014
```

```
;      OTHER INFORMATION:
US-08-414-657D-5


alignment_scores:
            Quality:   85.00              Length:    130
              Ratio:  1.308                Gaps:      7
 Percent Similarity:  50.000   Percent Identity: 27.692

alignment_block:
 US-09-060-609-2 x US-08-414-657D-5   ..

 Align seg 1/1  to: US-08-414-657D-5  from: 1  to: 1014

     101 SerAlaGlyGlyGluGluSerLeuLysCysGluAspLeuLysValGly.. 116
         ::::::|||   :::   |||||||||||||||         |||
     688 ACCACAGGACGACAAGCTTCCCTCAAATGTGAAGCCTCAGCGGTGCCTGC 737

     117 .........GlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130
                 ::::::         |||   :::|||||||:::|||:::
     738 ACCTGACTTTGAGTGGTACCGGGATGACACCAGGATAAACAGTGCAAACG 787

     130 lnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPhePro 146
                ::::::   |||:::               |||   :::
     788 GCCTTGAGATTAAGAGCACTGAG.........GGCCAGTCCTCCCTGACG 828

     147 AlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheTh 163
         :::   |||:::|||   ::::::     ||||||   |||      ::
     829 GTGACCAACGTCACT...GAGGAACACTACGGCAACTATACCTGTGTGGC 875

     163 rGlyAsnGluValGly.....................PhePheLysProI 173
         ::::|||:::::::|||                    :::|||:::|||
     876 TGCCAACAAGCTCGGCGTCACCAATGCCAGCCTAGTCCTTTTCAGACCCG 925

     173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
           |||   |||::::::||||||   |||   ::::::||||||   |||
     926 GGTCGGTGAGAGGAATCAACGGA...TCCATCAGTCTGGCCGTACCACTG 972

     190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
                        ||||||
     973 ..............TGGCTG............................... 978

     206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCys 219
             |||:::         ||||||   ::::::   |||
     979 ....CTGGCAGCGTCCCTGTTCTGCCTTCTCAGCAAATGT 1014

Seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-177-109A-1

;_documentation_block:
; Sequence 1, Application US/08177109A
; Patent No. 5869615
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/177,109A
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2483 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  116..2407
;      OTHER INFORMATION:  /note= "Product = Human Factor B"
US-08-177-109A-1


alignment_scores:
            Quality:   84.50              Length:    225
              Ratio:  0.790                Gaps:     13
 Percent Similarity:  47.556   Percent Identity: 23.111

alignment_block:
 US-09-060-609-2 x US-08-177-109A-1   ..

 Align seg 1/1  to: US-08-177-109A-1  from: 1  to: 2483

       2 HisIleLeuLysGlySerProAsnValIleProArgAlaHisGlyGlnLy 18
         |||         |||         ::::::::::|||
      74 CATCCTGCCCCAGGCCCAGCTTCTCTCCTGCCT.................. 106

      18 sAsnThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProPheL 35
             ::::::       |||::::::|||   |||
     107 ....TCCAACGCCATGGGGAGCAATCTCAGCCCC................. 136

      35 ysAsnLeuAlaLeuLeuProPheSerLeuProLeuLeuGlyGlyGly... 50
          :::|||:::|||:::||||||   |||   ||||||::::||||||
     137 ..CAACTCTGCCTGATGCCCTTTATCTTGGGCCTCTTGTCTGGAGGTGTG 184

      51 ................................GlySerGlySerGlyGl 56
                                        ||||||   |||   ||
     185 ACCACCACTCCATGGTCTTTGGCCCAGCCCCAGGGATCCTGCTCTCTGGA 234

      56 uLysValSerVal.................................SerL 62
         |   |||::::::                                 :::
     235 GGGGGTAGAGATCAAAGGCGGCTCCTTCCGACTTCTCCAAGAGGGCCAGG 284

      62 ysMetAlaAlaAlaTrpProSerGly......ProSerAlaProGluAla 76
          :::      :::  |||||||||      |||         ::::::
     285 CACTGGAGTACGTGTGTCCTTCTGGCTTCTACCCGTACCCTGTGCAGACA 334

      77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGl 93
            |||:::|||                                :::||||
     335 CGTACCTGCAGA.............................TCTACGGG 354

      93 yProTrpGlyAlaValAlaThrSer..................AlaGlyG 104
         |   |||:::::::::   |||:::                    :::
     355 GTCCTGGAGCACCCTGAAGACTCAAGACCAAAAGACTGTCAGGAAGGCAG 404

     104 lyGluGluSerLeuLysCys.........GluAspLeuLysValGlyGln 117
              :::::::::   |||         :::|||:::::   |||:::
     405 AGTGCAGAGCAATCCACTGTCCAAGACCACACGACTTCGAGAACGGGGAA 454

     118 TyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluProValAs 134
         |||       :::::|||       |||   :::   :::         ::
     455 TACTGGCCCCGGTCTCCCTACTACAATGTGAGTGATGAGATCTCTTTCCA 504
```

```
    134 nCys...ThrAsnTyrThrAlaHisValSerCysPheProAlaProAsnI 150
        :|||      :::|||||||   :::   |||:::            |||
    505 CTGCTATGACGGTTACACTCTCCGGGGCTCTGCC.............AATC 542

    150 leThrCysLysAspSerSer......GlyAsnGluThrHisPheThrGly 164
          ||||||:::   ::::::      ::::::::::|||   :::
    543 GCACCTGCCAAGTGAATGGCCGGTGGAGTGGGCAGACAGCGATCTGTGAC 592

    165 AsnGluValGlyPhePheLys............ProIleSerCysArgAs 177
        |||   :::|||:::   :::           ||||||:::   |||::
    593 AACGGAGCGGGGTACTGCTCCAACCCGGGCATCCCCATTGGCACAAGGAA 642

    177 nValAsnGlyTyrSerTyrLysVal 185
        :|||   |||   :::|||:::::
    643 GGTG...GGCAGCCAGTACCGCCTT 664

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-08-687-706-1

seq_documentation_block:
; Sequence 1, Application US/08687706
; Patent No. 5928892
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,706
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/177,109
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107 DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2483 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  116..2407
;      OTHER INFORMATION:  /note= "Product = Human Factor B"
US-08-687-706-1

alignment_scores:
```

```
            Quality:   84.50            Length:    225
              Ratio:  0.790              Gaps:     13
 Percent Similarity:  47.556   Percent Identity: 23.111

alignment_block:
 US-09-060-609-2 x US-08-687-706-1   ..

 Align seg 1/1  to: US-08-687-706-1  from: 1  to: 2483

      2 HisIleLeuLysGlySerProAsnValIleProArgAlaHisGlyGlnLy 18
        |||       |||     :::::::::|||
     74 CATCCTGCCCCAGGCCCAGCTTCTCTCCTGCCT.................. 106

     18 sAsnThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProPheL 35
          ::::::      |||:::::|||   |||
    107 ....TCCAACGCCATGGGGAGCAATCTCAGCCCC................ 136

     35 ysAsnLeuAlaLeuLeuProPheSerLeuProLeuLeuGlyGlyGly... 50
        :::|||:::|||:::||||||   |||   ||||||:::||||||
    137 ..CAACTCTGCCTGATGCCCTTTATCTTGGGCCTCTTGTCTGGAGGTGTG 184

     51 ..............................GlySerGlySerGlyGl 56
                                      ||||||   |||   ||
    185 ACCACCACTCCATGGTCTTTGGCCCAGCCCCAGGGATCCTGCTCTCTGGA 234

     56 uLysValSerVal..............................SerL 62
        |   |||::::::                            :::
    235 GGGGGTAGAGATCAAAGGCGGCTCCTTCCGACTTCTCCAAGAGGGCCAGG 284

     62 ysMetAlaAlaAlaTrpProSerGly......ProSerAlaProGluAla 76
        :::       :::   |||||||||       |||          ::::::
    285 CACTGGAGTACGTGTGTCCTTCTGGCTTCTACCCGTACCCTGTGCAGACA 334

     77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGl 93
          |||:::|||                                :::||||||
    335 CGTACCTGCAGA.............................TCTACGGG 354

     93 yProTrpGlyAlaValAlaThrSer.................AlaGlyG 104
        |   |||:::::::::   |||:::                    :::
    355 GTCCTGGAGCACCCTGAAGACTCAAGACCAAAAGACTGTCAGGAAGGCAG 404

    104 lyGluGluSerLeuLysCys.........GluAspLeuLysValGlyGln 117
          :::::::::   |||       :::|||::::::   |||:::
    405 AGTGCAGAGCAATCCACTGTCCAAGACCACACGACTTCGAGAACGGGGAA 454

    118 TyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluProValAs 134
        |||       ::::::|||       |||   :::   :::          ::
    455 TACTGGCCCCGGTCTCCCTACTACAATGTGAGTGATGAGATCTCTTTCCA 504

    134 nCys...ThrAsnTyrThrAlaHisValSerCysPheProAlaProAsnI 150
        :|||      :::||||||   :::   |||:::            |||
    505 CTGCTATGACGGTTACACTCTCCGGGGCTCTGCC.............AATC 542

    150 leThrCysLysAspSerSer......GlyAsnGluThrHisPheThrGly 164
          ||||||:::   ::::::      ::::::::::|||   :::
    543 GCACCTGCCAAGTGAATGGCCGGTGGAGTGGGCAGACAGCGATCTGTGAC 592

    165 AsnGluValGlyPhePheLys............ProIleSerCysArgAs 177
        |||   :::|||:::   :::           ||||||:::   |||::
    593 AACGGAGCGGGGTACTGCTCCAACCCGGGCATCCCCATTGGCACAAGGAA 642

    177 nValAsnGlyTyrSerTyrLysVal 185
        :|||   |||   :::|||:::::
    643 GGTG...GGCAGCCAGTACCGCCTT 664

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-121-057-2

seq_documentation_block:
; Sequence 2, Application US/08121057
; Patent No. 5484727
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
```

```
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,057
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-121-057-2


alignment_scores:
            Quality:   84.00              Length:    245
              Ratio:  0.764                Gaps:     13
 Percent Similarity:  44.898   Percent Identity: 24.082

alignment_block:
 US-09-060-609-2 x US-08-121-057-2    ..

 Align seg 1/1  to: US-08-121-057-2  from: 1  to: 4079

      72 SerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrp...Ph 87
         |||   |||:::::::::   |||:::|||:::   |||   |||   ::
    3248 TCCATTCCTAGGTCACTTGAAGCCAAACTGTTGGAAGTTCACTGGAGTCT 3297

      87 eValSerValThrThrGlyPro.........TrpGlyAlaValAlaThrS 101
         :|||   ::::::   |||            ||||||   |||
    3298 TGTACACTTAAGCAGAGGAGAACTTTTTTTGTGGGGCTGGGTG....... 3340

     101 erAlaGlyGlyGluGluSerLeuLysCysGlu.................. 111
              |||||||||:::   |||   :::|||
    3341 .....GGGGGAGAAGACCGACTAACAGCTGAAGTAATGACAGATTGTTGC 3385

     112 ............AspLeuLysValGlyGlnTyrIleCys........... 120
                    |||:::::::|||:::||||||||||
    3386 TGGGTCATATCAGCTTTATCCCTTGGTAATTATATCTGTTTTGTTTCTTG 3435

     121 ....Lys..AspProLysIleAsnAspAlaThrGlnGluProValAsnCy 135
                :::   :::||||||      ::::::   |||:::|||
    3436 ACTCTGTCCAATCAGAGAATAAACATCATAGTTTCTTGGCCACTGAATTA 3485

     135 sThrAsnTyr.....................ThrAlaHisValSerCysP 145
             ::::::                   |||:::   ::::::   |
    3486 GCCAAAACACTTAGGAAGAAATCACTTAAATACCTCTGGCTTAGAAATTT 3535

     145 heProAlaProAsnIleThrCys.........LysAspSerSerGlyAsn 158
         ||   :::      :::   |||         :::   :::   |||:::
    3536 TTTCATGCACACTGTTGGAATGTATGCTAATTGAACATGCAATTGGGGAA 3585

     159 GluThr.......................................... 160
         |||
    3586 GAAAAAATTAGAATGATTTTTGCTATTTCTAGTAGAAAGAAAATGTCTGT 3635

     161 .............HisPheThrGlyAsnGluValGlyPhePheLysProI 173
                        :::|||:::      |||   |||:::|||   :
    3636 TTTCCAAAGATAATGTTATACATCCTATTTTGTAATTTTTTTGAAAAAAG 3685

     173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
         :::::|||
    3686 TTCAATGT.......................................... 3693

     190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
         |||:::||||||:::                  |||||||||   :::
    3694 TCAGTTTTCCTTAGT.................TTTTACCTTGTTTTC.. 3723

     206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGlyS 223
          :::::::|||         ||||||   :::      |||   :::
    3724 .TCTATAGGTCAGTAT...TTCTGTGAAGCAAAAAGATGCCTTTTACCAT 3769

     223 erLeuIleAspPheIleLeuIleSerMetGlnIleValGlyProSerAsp 239
           :::::::::|||      |||   :::   |||   |||      :::
    3770 GAATTCTTGAGTTTACATCAATAATATTGTATATTAAGGGGATCAGAAGT 3819

     240 GlySerSerTyrIleIleAspTyrTyrGlyThrArgLeuThrArgLeuSe 256
           ::::::  |||   |||      |||   |||   |||
    3820 AGGAAGGAAAAAATAAGAGATAGCAGAGGAAAAAGAAAAACATTTCCTCT 3869

     256 rIleThrAsnGluThrPheArgLysThrGlnLeu 267
         ||||||:::|||::::::   |||   :::|||
    3870 TATAACTTCTGAAGTAATTTGTAAAAAAGATTTG 3903

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-509-187D-2

seq_documentation_block:
; Sequence 2, Application US/08509187D
; Patent No. 5834283
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Ta-Yuan and Chang, Catherine C.Y.
;    TITLE OF INVENTION:  ACYL Coenzyme A: Cholesterol Acyltransferase
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187D
;      FILING DATE:  31-JUL-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lamport Hammitte, Ann
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033cpdv
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-509-187D-2


alignment_scores:
           Quality:   84.00              Length:    245
             Ratio:  0.764                Gaps:     13
 Percent Similarity:  44.898   Percent Identity: 24.082

alignment_block:
 US-09-060-609-2 x US-08-509-187D-2   ..

 Align seg 1/1  to: US-08-509-187D-2  from: 1  to: 4079

      72 SerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrp...Ph 87
         |||   |||::::::::::   |||:::|||:::   |||   |||   ::
    3248 TCCATTCCTAGGTCACTTGAAGCCAAACTGTTGGAAGTTCACTGGAGTCT 3297

      87 eValSerValThrThrGlyPro.........TrpGlyAlaValAlaThrS 101
         :|||   ::::::   |||             ||||||   |||
    3298 TGTACACTTAAGCAGAGGAGAACTTTTTTTGTGGGGCTGGGTG....... 3340

     101 erAlaGlyGlyGluGluSerLeuLysCysGlu..................... 111
              |||||||||:::   |||   :::|||
    3341 .....GGGGGAGAAGACCGACTAACAGCTGAAGTAATGACAGATTGTTGC 3385

     112 ............AspLeuLysValGlyGlnTyrIleCys........... 120
                    |||::::::|||:::|||||||||
    3386 TGGGTCATATCAGCTTTATCCCTTGGTAATTATATCTGTTTTGTTTCTTG 3435

     121 ....Lys..AspProLysIleAsnAspAlaThrGlnGluProValAsnCy 135
                 :::   :::||||||      ::::::   |||:::|||
    3436 ACTCTGTCCAATCAGAGAATAAACATCATAGTTTCTTGGCCACTGAATTA 3485

     135 sThrAsnTyr.....................ThrAlaHisValSerCysP 145
             ::::::                      |||:::   ::::::   |
    3486 GCCAAAACACTTAGGAAGAAATCACTTAAATACCTCTGGCTTAGAAATTT 3535

     145 heProAlaProAsnIleThrCys.........LysAspSerSerGlyAsn 158
         ||   :::      :::   |||        :::   :::   |||:::
    3536 TTTCATGCACACTGTTGGAATGTATGCTAATTGAACATGCAATTGGGGAA 3585

     159 GluThr.......................................... 160
         |||
    3586 GAAAAAATTAGAATGATTTTTGCTATTTCTAGTAGAAAGAAAATGTCTGT 3635

     161 .............HisPheThrGlyAsnGluValGlyPhePheLysProI 173
                     :::|||:::      |||   |||:::|||   :
    3636 TTTCCAAAGATAATGTTATACATCCTATTTTGTAATTTTTTTGAAAAAAG 3685

     173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
         :::::|||
    3686 TTCAATGT.......................................... 3693

     190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
         |||:::||||||:::                |||||||||   :::
    3694 TCAGTTTTCCTTAGT.................TTTTACCTTGTTTTC.. 3723

     206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGlyS 223
          ::::::|||         ||||||   :::      |||   :::
    3724 .TCTATAGGTCAGTAT...TTCTGTGAAGCAAAAAGATGCCTTTTACCAT 3769

     223 erLeuIleAspPheIleLeuIleSerMetGlnIleValGlyProSerAsp 239
           ::::::::::|||      |||   :::   |||   |||      :::
    3770 GAATTCTTGAGTTTACATCAATAATATTGTATATTAAGGGGATCAGAAGT 3819

     240 GlySerSerTyrIleIleAspTyrTyrGlyThrArgLeuThrArgLeuSe 256
          ::::::   |||   |||      |||   |||   |||
    3820 AGGAAGGAAAAAATAAGAGATAGCAGAGGAAAAAGAAAAACATTTCCTCT 3869

     256 rIleThrAsnGluThrPheArgLysThrGlnLeu 267
          ||||||:::|||::::::   |||   :::|||


    3870 TATAACTTCTGAAGTAATTTGTAAAAAAGATTTG 3903

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:US-09-121-396-2

seq_documentation_block:
; Sequence 2, Application US/09121396
; Patent No. 5968749
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,396
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187
;      FILING DATE:  07/31/95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CPDV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-121-396-2


alignment_scores:
           Quality:   84.00              Length:    245
             Ratio:  0.764                Gaps:     13
 Percent Similarity:  44.898   Percent Identity: 24.082

alignment_block:
 US-09-060-609-2 x US-09-121-396-2   ..

 Align seg 1/1  to: US-09-121-396-2  from: 1  to: 4079

      72 SerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrp...Ph 87
         |||   |||::::::::::   |||:::|||:::   |||   |||   ::
    3248 TCCATTCCTAGGTCACTTGAAGCCAAACTGTTGGAAGTTCACTGGAGTCT 3297

      87 eValSerValThrThrGlyPro.........TrpGlyAlaValAlaThrS 101
         :|||   ::::::   |||             ||||||   |||
    3298 TGTACACTTAAGCAGAGGAGAACTTTTTTTGTGGGGCTGGGTG....... 3340

     101 erAlaGlyGlyGluGluSerLeuLysCysGlu..................... 111
              |||||||||:::   |||   :::|||
    3341 .....GGGGGAGAAGACCGACTAACAGCTGAAGTAATGACAGATTGTTGC 3385

     112 ............AspLeuLysValGlyGlnTyrIleCys........... 120
                    |||::::::|||:::|||||||||
```

```
 3386 TGGGTCATATCAGCTTTATCCCTTGGTAATTATATCTGTTTTGTTTCTTG 3435

  121 ....Lys..AspProLysIleAsnAspAlaThrGlnGluProValAsnCy 135
              :::   :::||||||      ::::::   |||:::|||
 3436 ACTCTGTCCAATCAGAGAATAAACATCATAGTTTCTTGGCCACTGAATTA 3485

  135 sThrAsnTyr.....................ThrAlaHisValSerCysP 145
        ::::::                       |||:::   ::::::   |
 3486 GCCAAAACACTTAGGAAGAAATCACTTAAATACCTCTGGCTTAGAAATTT 3535

  145 heProAlaProAsnIleThrCys.........LysAspSerSerGlyAsn 158
      ||   :::      :::   |||         :::   :::   |||:::
 3536 TTTCATGCACACTGTTGGAATGTATGCTAATTGAACATGCAATTGGGGAA 3585

  159 GluThr.......................................... 160
      |||
 3586 GAAAAAATTAGAATGATTTTTGCTATTTCTAGTAGAAAGAAAATGTCTGT 3635

  161 .............HisPheThrGlyAsnGluValGlyPhePheLysProI 173
                     :::|||:::      |||   |||:::|||   :
 3636 TTTCCAAAGATAATGTTATACATCCTATTTTGTAATTTTTTTGAAAAAAG 3685

  173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
      :::::|||
 3686 TTCAATGT.......................................... 3693

  190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
      |||:::||||||:::                |||||||||  :::
 3694 TCAGTTTTCCTTAGT.................TTTTACCTTGTTTTC.. 3723

  206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGlyS 223
       ::::::|||         ||||||   :::      |||   :::
 3724 .TCTATAGGTCAGTAT...TTCTGTGAAGCAAAAAGATGCCTTTTACCAT 3769

  223 erLeuIleAspPheIleLeuIleSerMetGlnIleValGlyProSerAsp 239
        :::::::::|||      |||   :::   |||   |||      :::
 3770 GAATTCTTGAGTTTACATCAATAATATTGTATATTAAGGGGATCAGAAGT 3819

  240 GlySerSerTyrIleIleAspTyrTyrGlyThrArgLeuThrArgLeuSe 256
        ::::::   |||   |||      |||   |||   |||
 3820 AGGAAGGAAAAAATAAGAGATAGCAGAGGAAAAAGAAAAACATTTCCTCT 3869

  256 rIleThrAsnGluThrPheArgLysThrGlnLeu 267
       ||||||:::|||::::::   |||   :::|||
 3870 TATAACTTCTGAAGTAATTTGTAAAAAAGATTTG 3903

seq_name: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:PCT-US93-09704A-2

seq_documentation_block:
;  Sequence 2, Application PC/TUS9309704A
;   GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;     TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;     NUMBER OF SEQUENCES:  9
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII Text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/09704A
;       FILING DATE:  October 12, 1993
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  U.S. SER. NO. 959,950
;       FILING DATE:  October 14, 1992
;       APPLICATION NUMBER:  U.S. SER. NO. 121,057
;       FILING DATE:  September 10, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  LAMPORT HAMMITTE, ANN.
;       REGISTRATION NUMBER:  34,858
;       REFERENCE/DOCKET NUMBER:  DCI-033CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-2700
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4079 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
PCT-US93-09704A-2


alignment_scores:
            Quality:   84.00              Length:    245
              Ratio:  0.764                Gaps:     13
 Percent Similarity:  44.898   Percent Identity: 24.082

alignment_block:
 US-09-060-609-2 x PCT-US93-09704A-2   ..

 Align seg 1/1  to: PCT-US93-09704A-2  from: 1  to: 4079

   72 SerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrp...Ph 87
      |||   |||:::::::::   |||:::|||:::   |||   |||   ::
 3248 TCCATTCCTAGGTCACTTGAAGCCAAACTGTTGGAAGTTCACTGGAGTCT 3297

   87 eValSerValThrThrGlyPro.........TrpGlyAlaValAlaThrS 101
      :|||   ::::::   |||             ||||||   |||
 3298 TGTACACTTAAGCAGAGGAGAACTTTTTTTGTGGGGCTGGGTG....... 3340

  101 erAlaGlyGlyGluGluSerLeuLysCysGlu..................... 111
          ||||||||||:::   |||   :::|||
 3341 .....GGGGGAGAAGACCGACTAACAGCTGAAGTAATGACAGATTGTTGC 3385

  112 ............AspLeuLysValGlyGlnTyrIleCys........... 120
                    |||:::::|||:::||||||||||
 3386 TGGGTCATATCAGCTTTATCCCTTGGTAATTATATCTGTTTTGTTTCTTG 3435

  121 ....Lys..AspProLysIleAsnAspAlaThrGlnGluProValAsnCy 135
              :::   :::||||||      ::::::   |||:::|||
 3436 ACTCTGTCCAATCAGAGAATAAACATCATAGTTTCTTGGCCACTGAATTA 3485

  135 sThrAsnTyr.....................ThrAlaHisValSerCysP 145
        ::::::                       |||:::   ::::::   |
 3486 GCCAAAACACTTAGGAAGAAATCACTTAAATACCTCTGGCTTAGAAATTT 3535

  145 heProAlaProAsnIleThrCys.........LysAspSerSerGlyAsn 158
      ||   :::      :::   |||         :::   :::   |||:::
 3536 TTTCATGCACACTGTTGGAATGTATGCTAATTGAACATGCAATTGGGGAA 3585

  159 GluThr.......................................... 160
      |||
 3586 GAAAAAATTAGAATGATTTTTGCTATTTCTAGTAGAAAGAAAATGTCTGT 3635

  161 .............HisPheThrGlyAsnGluValGlyPhePheLysProI 173
                     :::|||:::      |||   |||:::|||   :
 3636 TTTCCAAAGATAATGTTATACATCCTATTTTGTAATTTTTTTGAAAAAAG 3685

  173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
      :::::|||
 3686 TTCAATGT.......................................... 3693

  190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
      |||:::||||||:::                |||||||||  :::
 3694 TCAGTTTTCCTTAGT.................TTTTACCTTGTTTTC.. 3723

  206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGlyS 223
       ::::::|||         ||||||   :::      |||   :::
 3724 .TCTATAGGTCAGTAT...TTCTGTGAAGCAAAAAGATGCCTTTTACCAT 3769

  223 erLeuIleAspPheIleLeuIleSerMetGlnIleValGlyProSerAsp 239
        :::::::::|||      |||   :::   |||   |||      :::
```

```
    3770 GAATTCTTGAGTTTACATCAATAATATTGTATATTAAGGGGATCAGAAGT 3819

     240 GlySerSerTyrIleIleAspTyrTyrGlyThrArgLeuThrArgLeuSe 256
         ::::::   |||   |||      |||   |||   |||
    3820 AGGAAGGAAAAAATAAGAGATAGCAGAGGAAAAAGAAAAACATTTCCTCT 3869

     256 rIleThrAsnGluThrPheArgLysThrGlnLeu 267
         ||||||:::|||::::::   |||   :::|||
    3870 TATAACTTCTGAAGTAATTTGTAAAAAAGATTTG 3903

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-276-967-1

seq_documentation_block:
; Sequence 1, Application US/08276967
; Patent No. 5851817
;  GENERAL INFORMATION:
;    APPLICANT:  Hardy, Daniel M.
;    APPLICANT:  Garbers, David L.
;    TITLE OF INVENTION:  Species-Specific Egg-Binding Proteins of
;    TITLE OF INVENTION:  Sperm
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,967
;      FILING DATE:  Submitted Herewith
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  UTSD:418\KIT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7785 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-276-967-1


alignment_scores:
            Quality:   83.00              Length:    278
              Ratio:  0.709                Gaps:     16
 Percent Similarity:  42.086   Percent Identity: 23.022

alignment_block:
 US-09-060-609-2 x US-08-276-967-1   ..

 Align seg 1/1  to: US-08-276-967-1  from: 1  to: 7785

      13 ArgAlaHisGlyGlnLysAsnThrArgArgAspGlyThrGlyLeuTyrPr 29
         ||||||      |||      |||   :::   :::   :::      ||
    2360 CGAGCTGCCCCCCAAATGCCCACTTTGAACGCTGCGCCTGCCCAGTGTCC 2409

      29 oMetArgGlyProPheLysAsnLeuAla..................... 38
         |   |||:::|||      :::::::::
    2410 TGCCAGAGCCCCACACCCAACTGTGAGCTCTTCTGCAAGCCCGGCTGTGT 2459

      39 .....LeuLeuProPhe.....SerLeu 44
              :::|||   |||        |||
    2460 CTGTGATCCTGGCTTTTTATTCAGTGGCTCCCACTGCGTCAACGCCTCTT 2509

      45 ProLeuLeuGlyGlyGlyGlySerGlySerGlyGluLysValSerValSe 61
         |||::::::                        :::|||:::::
    2510 CCTGTGATT.....GCTTCTACAAC 2529

      61 rLysMetAlaAlaAlaTrpProSerGlyProSerAlaProGluAlaValT 78
         :  ::::::   :::|||   :::      |||||||||         |
    2530 GACAATTACTATAAGCTGGGGACAGATTGGTTCAGCCCCA.....A 2570

      78 hrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGlyPro 94
         |||||:::   :::                  :::|||:::|||||||
    2571 CTGCACAGAACATT.....GCCACTGCCGGCCCA 2599

      95 .....TrpGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLe 108
              |||:::|||      :::|||||||||      :::
    2600 GCAGTCGGATGGAGTGCCAGACCTTCAAGTGCGGGACACACACAGTG... 2646

     108 uLysCysGluAspLeuLysValGlyGlnTyrIleCysLys..... 121
         |||:::   ||||||   |||||||||   |||
    2647 ....TGCCAG...CTGAAGAATGGCCAGTACGGCTGCCACCCCTATGGCA 2689

     122 .....AspPro..... 123
                 ||||||
    2690 GTGCCACCTGCTCTGTCTACGGAGACCCTCACTACCTCACCTTCGACGGG 2739

     124 .....LysIleAsnAspAlaThrGlnGluProVa 133
                      |||   :::        :::|||
    2740 AGGCGCTTTAACTTCATGGGCAAGTGCACCTACATCTTGGCCCAACCC.. 2787

     133 lAsnCysThrAsnTyrThrAlaHis..... 141
         |||   |||   |||   |||
    2788 ....TGTGGCAACTTGACAGAGCACTTCTTCAGGGTGCTGGTGAAGAAGG 2833

     142 .....ValSerCysPheProAlaProAsnIleThr 151
                      ||||||||||:::        :::|||
    2834 AGGAGCGAGGACAGGAGGGCGTGTCCTGCCTAAGCAAGGTCTACGTGACT 2883

     152 CysLysAspSerSer.....GlyAsnGluThrHisPheTh 163
              :::|||:::      |||::::::|||
    2884 CTGCCTGAAAGCACCGTCACTCTGCTCAAGGGCAGACACACGCTGGTCGG 2933

     163 rGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVal.... 178
          |||::::::|||      :::      |||      |||:::|||
    2934 AGGTCAGCGAGTC...ACCCTCCCAGCCATACCTTCTAGAGGTGTCTTCC 2980

     179 .......AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPhe 192
                :::|||        ::::::   :::|||::::::|||
    2981 TGGCTCCCAGTGGGCGATTTGTGGAGCTGCAGACGGCGTTCGGTCTGCGG 3030

     193 LeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGl 209
         :::   |||   |||   ::::::::::::::::::   |||:::
    3031 GTGAGATGGGATGGTGACCAGCAGCTGTTTGTGAGTGTGCCCAGCACCTT 3080

     209 yLeuLeuLysPheCysThrValGlyPheCysGly 220
              |||:::|||      |||:::||||||
    3081 CTCTGGCAAACTCTGT......GGTCTCTGTGGC 3108

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-828-832-2

seq_documentation_block:
; Sequence 2, Application US/08828832
; Patent No. 5827711
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
```

```
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/828,832
;       FILING DATE:  Herewith
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0250 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  649 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  Consensus
;       CLONE:  2454416
US-08-828-832-2


alignment_scores:
            Quality:   82.00          Length:    193
              Ratio:  0.845            Gaps:     11
 Percent Similarity:  50.259   Percent Identity: 24.870

alignment_block:
 US-09-060-609-2 x US-08-828-832-2   ..

 Align seq 1/1  to: US-08-828-832-2  from: 1  to: 649

     82 ValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyAlaVa 98
        ::::::|||||||||   :::|||                  |||||
     66 ATGGCGGTTCTCTGGAGGCTGAGT.....................GCCGT 94

     98 lAlaThrSerAlaGlyGly.Glu.........GluSerLeuLysCysGlu 111
        |:::   :::   ||||||  |||           :::|||:::   :::
     95 TTGCGGTGCCCTAGGAGGCCGAGCTCTGTTGCTTCGAACTCCAGTGGTCA 144

    112 AspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAl 128
        ||||||   ::::::|||::::::   :::            :::|||
    145 GACCTGCTCATATCTCAGCATTTCTTCAGG............ACCGACCT 182

    128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
         :::|||:::   |||:::|||::::::||||||:::|||   :::   :
    183 ATCCCAGAATGGTGTGGAGTGCAGCACATACACTTGTCACCGAGCCACCA 232

    145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
        ::   ||||||::::::                    :::   |||
    233 TTCTGGCTCCAAGGCTGCA.....................TCTCTCCAC 260

    162 PheThrGlyAsnGluVal.............................. 167
        :::|||::::::::::|||
    261 TGGACTAGCGAGAGGGTTGTCAGTGTTTGCTCCTGGGTCTGCTTCCGGC 310

    168 .GlyPheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrL 184
         ::::::::::::|||   |||:::   :::               |||:
    311 TGCTTATTTGAATCCTTGCTCTGCGATGGAC..............TATT 345

    184 ysValAlaValAlaLeuSerLeuPheLeuGlyTrp...LeuGly..... 197
        :::::|||:::||||||:::|||          |||   ||||||
    346 CCCTGGCTGCAGCCCTCACTCTTCATGGTCACTGGGGCCTTGGACAAGTT 395

    197 ..... 197

    396 GTTACTGACTATGTTCATGGGGATGCCTTGCAGAAAGCTGCCAAGGCAGG 445

    198 .....AlaAspArgPheTyrLeuGlyT 205
                          |||         |||::::::|
    446 GCTTTTGGCACTTTCAGCTTTAACCTTTGCTGGGCTTTGCTATTTCAACT 495

    205 yrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPhe.CysGlyIl 221
        ||      :::|||:::   |||      :::|||:::::: ||||||
    496 ATCACGATGTGGGCATCTGCAAA......GCTGTTGCCATGCTGTGGAAG 539

    221 eGlySerLeuIleAspPheIleLeu 229
         ::::::|||:::||||||||||||
    540 CTCTGACCTTTTTGACTTCATACTT 564
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 07:47:33 ; Search time 13.25 Seconds
(without alignments)
957.626 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 269
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: OLIGO

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	3.0	1520	1	TVFFA	protein-tyrosine k
2	8	3.0	358	2	S23383	protein kinase (EC
3	8	3.0	376	2	S22745	serine/threonine p
4	8	3.0	288	2	T05954	transcription fact
5	8	3.0	473	2	S70357	forkhead transcrip
6	8	3.0	256	2	T03371	glycine-rich prote
7	8	3.0	252	2	T03160	capsid protein - a
8	8	3.0	427	2	H64938	hypothetical prote
9	8	3.0	529	2	S76167	hypothetical prote
10	8	3.0	2946	2	T00867	hypothetical prote
11	8	3.0	322	2	T04595	hypothetical prote
12	8	3.0	207	2	T07381	glycine-rich prote
13	8	3.0	548	2	T16082	hypothetical prote
14	8	3.0	173	2	A47303	FTZ-F1 steroid rec
15	8	3.0	1174	2	A40853	potassium channel
16	8	3.0	613	2	I39295	X-linked PEST-cont
17	8	3.0	598	2	I38495	X-linked PEST-cont
18	7	2.6	419	1	JQ0346	ubiquinol--cytochr
19	7	2.6	451	1	JE0240	LIM kinase (EC 2.7
20	7	2.6	551	1	NRECE3	colicin E3 (EC 3.1
21	7	2.6	576	1	S22453	colicin E7 (EC 3.1
22	7	2.6	581	1	NDECE2	colicin E2 (EC 3.1
23	7	2.6	482	1	ZPBY	mitochondrial proc
24	7	2.6	577	1	A36442	mitochondrial proc
25	7	2.6	623	1	S33167	pointed protein, s
26	7	2.6	718	1	S33168	pointed protein, s
27	7	2.6	1959	1	AGRT	agrin - rat
28	7	2.6	166	1	KRBO2B	keratin, 68K type
29	7	2.6	183	1	KNRZG2	glycine-rich cell
30	7	2.6	835	1	A45596	trypomastigote-spe
31	7	2.6	511	1	IKECBB	colicin B - Escher
32	7	2.6	424	1	Z3BPFD	coat protein A pre
33	7	2.6	424	1	Z3BPM3	coat protein A pre
34	7	2.6	424	1	Z3BPF1	coat protein A pre
35	7	2.6	248	2	F71538	probable oxoacyl (
36	7	2.6	679	2	C71007	probable formate d
37	7	2.6	399	2	I38901	JNK-activating pro
38	7	2.6	638	2	I78846	LIM protein kinase
39	7	2.6	617	2	I78847	LIM protein kinase
40	7	2.6	638	2	JC5813	LIM-kinase (EC 2.7
41	7	2.6	617	2	JC5814	LIM-kinase (EC 2.7
42	7	2.6	163	2	S11399	PTS fructose-speci
43	7	2.6	173	2	G71916	inorganic pyrophos
44	7	2.6	173	2	D64597	inorganic pyrophos
45	7	2.6	301	2	G72698	hypothetical prote

ALIGNMENTS

RESULT 1
TVFFA
protein-tyrosine kinase (EC 2.7.1.112) abl - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 15-Nov-1984 #sequence_revision 30-Sep-1989 #text_change 11-Jun-1999
C;Accession: A28128; A00628
R;Henkemeyer, M.J.; Bennett, R.L.; Gertler, F.B.; Hoffmann, F.M.
Mol. Cell. Biol. 8, 843-853, 1988
A;Title: DNA sequence, structure, and tyrosine kinase activity of the Drosophila mela
A;Reference number: A28128; MUID:88174728
A;Accession: A28128
A;Molecule type: DNA
A;Residues: 1-1520 <HEN>
A;Cross-references: GB:M19692; GB:M18402; NID:g158598; PIDN:AAA28934.1; PID:g158600
R;Hoffmann, F.M.; Fresco, L.D.; Hoffman-Falk, H.; Shilo, B.Z.
Cell 35, 393-401, 1983
A;Title: Nucleotide sequences of the Drosophila src and abl homologs: conservation an
A;Reference number: A00628; MUID:84082064
A;Accession: A00628
A;Molecule type: DNA
A;Residues: 'A',375,'AQ',378-644,'VGDV' <HOF>
A;Cross-references: GB:K01042; NID:g157175; PIDN:AAA28443.1; PID:g157176
C;Genetics:
A;Gene: abl
A;Cross-references: FlyBase:FBgn0000017
A;Introns: 112/1; 130/1; 310/2; 449/1; 644/3; 735/1; 775/2; 805/1; 1350/1
C;Superfamily: Drosophila protein-tyrosine kinase abl; protein kinase homology; SH2 h
C;Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; transformin
F;211-260/Domain: SH3 homology <SH3>
F;271-363/Domain: SH2 homology <SH2>
F;386-646/Domain: protein kinase homology <KIN>
F;394-402/Region: protein kinase ATP-binding motif
F;417/Active site: Lys #status predicted

Query Match 3.0%; Score 8; DB 1; Length 1520;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     48 GGGGSGSG 55
          ||||||||
Db     78 GGGGSGSG 85
```

RESULT 2
S23383
protein kinase (EC 2.7.1.37) cdc2-related KKIALRE - human
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 18-Jun-1999
C;Accession: S23383; S22744
R;Meyerson, M.; Enders, G.H.; Wu, C.L.; Su, L.K.; Gorka, C.; Nelson, C.; Harlow, E.;
EMBO J. 11, 2909-2917, 1992
A;Title: A family of human cdc2-related protein kinases.
A;Reference number: S23382; MUID:92347325
A;Accession: S23383
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-358 <MEY>

A;Cross-references: EMBL:X66358; NID:g36614; PIDN:CAA47002.1; PID:g36615
C;Superfamily: kinase-related transforming protein; protein kinase homology
C;Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
F;3-278/Domain: protein kinase homology <KIN>
F;11-19/Region: protein kinase ATP-binding motif
F;34,52,127,129/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 3.0%; Score 8; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 YPALGLLK 212
||||||||
Db 259 YPALGLLK 266

RESULT 3
S22745
[illegible]ne/threonine protein kinase KKIALRE (EC 2.7.1.-) - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 19-Dec-1997
C;Accession: S22745
R;Meyerson, M.L.
submitted to the EMBL Data Library, May 1992
A;Reference number: S22743
A;Accession: S22745
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-376 <MEY>
A;Cross-references: EMBL:X66359
C;Genetics:
A;Introns: 152/3; 170/3
C;Superfamily: kinase-related transforming protein; protein kinase homology
C;Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
F;3-296/Domain: protein kinase homology <KIN>
F;11-19/Region: protein kinase ATP-binding motif

Query Match 3.0%; Score 8; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 YPALGLLK 212
||||||||
Db 277 YPALGLLK 284

[illegible]ULT 4
T05954
transcription factor myb4 - barley (fragment)
C;Species: Hordeum vulgare (barley)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999
C;Accession: T05954
R;Gubler, F.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z15470
A;Accession: T05954
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-288 <GUB>
A;Cross-references: EMBL:X99973; NID:e1251306; PIDN:CAA68235.1; PID:e258527
A;Experimental source: cv. Himalaya, endosperm
C;Genetics:
A;Gene: myb4
C;Superfamily: Arabidopsis 28K leaf-specific myb-related protein; myb DNA-binding repeat
C;Keywords: DNA binding; transcription regulation
F;60-110/Domain: myb DNA-binding repeat homology <MYB>

Query Match 3.0%; Score 8; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 265 GGGGSGSG 272

RESULT 5
S70357
forkhead transcription factor HNF-3 alpha - human
N;Alternate names: hepatocyte nuclear factor-3 alpha
C;Species: Homo sapiens (man)
C;Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 16-Jul-1999
C;Accession: S70357
R;Bingle, C.D.; Gowan, S.
Biochim. Biophys. Acta 1307, 17-20, 1996
A;Title: Molecular cloning of the forkhead transcription factor HNF-3-alpha from a hu
A;Reference number: S70357; MUID:96254057
A;Accession: S70357
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-473 <BIN>
A;Cross-references: EMBL:U39840; NID:g1066121; PIDN:AAB06493.1; PID:g1066122
C;Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
F;171-262/Domain: fork head DNA-binding domain homology <FHD>

Query Match 3.0%; Score 8; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 277 GGGGSGSG 284

RESULT 6
T03371
glycine-rich protein grp3 - maize
C;Species: Zea mays (maize)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 20-Aug-1999
C;Accession: T03371
R;Goddemeier, M.L.; Wulff, D.; Feix, G.
Plant Mol. Biol. 36, 799-802, 1998
A;Title: Root-specific expression of a Zea mays gene encoding a novel glycine-rich pr
A;Reference number: Z14904; MUID:98187261
A;Accession: T03371
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-256 <GOD>
A;Cross-references: EMBL:Y07781; NID:e1283638; PIDN:CAA69104.1; PID:e265498
A;Experimental source: strain DK105; root
C;Genetics:
A;Gene: grp3
C;Superfamily: glycine-rich cell wall structural protein 1

Query Match 3.0%; Score 8; DB 2; Length 256;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 242 GGGGSGSG 249

RESULT 7
T03160
capsid protein - alcelaphine herpesvirus 1
C;Species: alcelaphine herpesvirus 1
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 07-May-1999
C;Accession: T03160
R;Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. 71, 6517-6525, 1997

A;Title: Primary structure of the alcelaphine herpesvirus 1 genome.
A;Reference number: Z14840; MUID:97404659
A;Accession: T03160
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-252 <ENS>
A;Cross-references: EMBL:AF005370; NID:g2337967; PID:g2338028
C;Keywords: capsid protein

Query Match 3.0%; Score 8; DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 96 GGGGSGSG 103

[redacted]LT 8
938
hypothetical protein b1784 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 14-Nov-1997
C;Accession: H64938
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: H64938
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-427 <BLAT>
A;Cross-references: GB:AE000273; GB:U00096; NID:g1788078; PID:g1788084; UWGP:b1784
A;Experimental source: strain K-12, substrain MG1655

Query Match 3.0%; Score 8; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 94 GGGGSGSG 101

[redacted]LT 9
167
Hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C;Accession: S76167
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76167
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-529 <KAN>
A;Cross-references: EMBL:D90914; GB:AB001339; NID:g1653477; PID:d1019159; PID:g1653513
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 3.0%; Score 8; DB 2; Length 529;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 SLPLLGGG 50
||||||||
Db 260 SLPLLGGG 267

RESULT 10
T00867
hypothetical protein F17K2.7 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 14-May-1999
C;Accession: T00867
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, March 1998
A;Description: Arabidopsis thaliana chromosome II BAC F17K2 genomic sequence.
A;Reference number: Z14207
A;Accession: T00867
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2946 <ROU>
A;Cross-references: EMBL:AC003680; NID:g2979540; PID:g2979554
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 2
A;Introns: 652/1; 913/3; 1165/3; 1210/3; 1245/3; 1283/3; 1309/2; 1336/3; 1438/3; 1638
; 2449/2; 2481/3; 2643/1; 2890/3; 2931/3
A;Note: F17K2.7

Query Match 3.0%; Score 8; DB 2; Length 2946;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 LGGGGSGS 54
||||||||
Db 1507 LGGGGSGS 1514

RESULT 11
T04595
hypothetical protein F23E13.150 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04595
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse,
submitted to the Protein Sequence Database, March 1998
A;Reference number: Z15378
A;Accession: T04595
A;Molecule type: DNA
A;Residues: 1-322 <BEV>
A;Cross-references: EMBL:AL022141
A;Experimental source: cultivar Columbia; BAC clone F23E13
C;Genetics:
A;Map position: 4
A;Introns: 189/1
A;Note: F23E13.150

Query Match 3.0%; Score 8; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 149 GGGGSGSG 156

RESULT 12
T07381
glycine-rich protein Tfm5 - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 14-May-1999
C;Accession: T07381
R;Santino, C.G.; Stanford, G.L.; Conner, T.W.
Plant Mol. Biol. 33, 405-416, 1997

A;Title: Developmental and transgenic analysis of two tomato fruit enhanced genes.
A;Reference number: Z16000
A;Accession: T07381
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-207 <SAN>
A;Cross-references: EMBL:X95262; NID:g1166449; PID:e219171
A;Experimental source: cultivar UC82b; fruit
C;Genetics:
A;Gene: Tfm5

Query Match 3.0%; Score 8; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 45 GGGGSGSG 52

RESULT 13
T16082
hypothetical protein F16H11.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16082
R;Wu, X.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid F16H11.
A;Reference number: Z18458
A;Accession: T16082
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-548 <WUX>
A;Cross-references: EMBL:U55376; NID:g1280130; PID:g1280133; PIDN:AAA98005.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone F16H11
C;Genetics:
A;Gene: CESP:F16H11.5
A;Map position: 10
A;Introns: 71/1; 93/1; 187/3; 281/3; 316/3; 420/2; 460/1; 513/3; 545/3

Query Match 3.0%; Score 8; DB 2; Length 548;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 KDSSGNET 160
||||||||
Db 109 KDSSGNET 116

RESULT 14
A47303
FTZ-F1 steroid receptor beta isoform - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 30-May-1997
C;Accession: A47303
R;Lavorgna, G.; Karim, F.D.; Thummel, C.S.; Wu, C.
Proc. Natl. Acad. Sci. U.S.A. 90, 3004-3008, 1993
A;Title: Potential role for a FTZ-F1 steroid receptor superfamily member in the control
A;Reference number: A47303; MUID:93219413
A;Accession: A47303
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-173 <LAV>
A;Note: sequence extracted from NCBI backbone (NCBIN:128795, NCBIP:128796)
C;Genetics:
A;Gene: FlyBase:ftz-f1
A;Cross-references: FlyBase:FBgn0001078
C;Keywords: steroid hormone receptor

Query Match 3.0%; Score 8; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 96 GGGGSGSG 103

RESULT 15
A40853
potassium channel protein eag - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change 24-Sep-1998
C;Accession: A40853
R;Warmke, J.; Drysdale, R.; Ganetzky, B.
Science 252, 1560-1562, 1991
A;Title: A distinct potassium channel polypeptide encoded by the Drosophila eag locus
A;Reference number: A40853; MUID:91262635
A;Accession: A40853
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1174 <WAR>
A;Cross-references: GB:M61157; NID:g157311; PID:g157312
C;Genetics:
A;Gene: FlyBase:eag
A;Cross-references: FlyBase:FBgn0000535
C;Keywords: transmembrane protein

Query Match 3.0%; Score 8; DB 2; Length 1174;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGSG 55
||||||||
Db 1125 GGGGSGSG 1132

Search completed: February 11, 2000, 08:02:52
Job time: 919 sec

```
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=BRAIN;
RX   MEDLINE; 95220355.
RA   SCHRODER S., MORRIS S.A., KNORR R., PLESSMANN U., WEBER K.,
RA   VINH N.G., UNGEWICKELL E.;
RT   "Primary structure of the neuronal clathrin-associated protein
RT   auxilin and its expression in bacteria.";
RL   Eur. J. Biochem. 228:297-304(1995).
CC   -!- FUNCTION: BINDS CLATHRIN HEAVY CHAIN AND PROMOTES ITS ASSEMBLY
CC       INTO REGULAR CAGES.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- PTM: TARGET FOR COAT-ASSOCIATED CASEIN KINASE II IN VITRO.
CC   -!- SIMILARITY: CONTAINS A TENSIN DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09237; AAA79037.1; -.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; UNKNOWN_1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Coat protein; Repeat; Phosphorylation.
FT   DOMAIN       33     44       3 X 4 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT       33     36       1.
FT   REPEAT       37     40       2.
FT   REPEAT       41     44       3.
FT   DOMAIN       60    344       TENSIN.
FT   DOMAIN      406    414       SH3-BINDING (POTENTIAL).
FT   DOMAIN      463    757       PRO-RICH.
FT   DOMAIN      526    529       POLY-GLY.
FT   DOMAIN      846    910       DNAJ-LIKE.
SQ   SEQUENCE   910 AA;  99512 MW;  CD5B90CA CRC32;


  Query Match             5.5%;  Score 79.5;  DB 1;  Length 910;
  Best Local Similarity   24.3%;  Pred. No. 26;
  Matches   46;  Conservative   22;  Mismatches   86;  Indels   35;  Gaps    8;

Qy      33 PFKNLALLPFSLPLLGGGGSGSGEKVSVSKMAAAW-PSGPSAPEAVTAR---------- 81
           |  |  ||     | ||||:    :   | :   : ||||:: ::   |
Db     514 PPSNSELLS---DLFGGGGAAGPVQSGQSGVDDVFHPSGPTSTQSTPRRSATSTSASPTL 570

Qy      81 LVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCTNYTA 140
            ||       |  |:|| :  :|       :|| :| ::       :|   :|    : |
Db     571 RVG----EGATFDPFGAPSKPSG-------QDL-LGSFLNTASASSDPFLQPTRSPSPTV 618

Qy     141 HVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAV------ALSLF-- 193
           | |  || ||    |   : |      |     :  :  | |:           |  |
Db     619 HASSTPAVNIQPDVSGAWDWHTKPGGFGMGSKSAATSPTGSSHGTPTHQNKPQTLDPFAD 678

Qy     193 LGWLGADRF 201
           || ||   |
Db     679 LGTLGGSSF 687


Search completed: February 11, 2000, 04:29:59
Job time: 432 sec
```

```
DR   MIM; 146731; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
DR   PFAM; PF00086; thyroglobulin_1; 1.
DR   PFAM; PF00219; IGFBP; 1.
KW   Growth factor binding; Signal.
FT   SIGNAL        1     39
FT   CHAIN        40    328       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN 2.
FT   DOMAIN      260    309       THYROGLOBULIN TYPE I.
FT   SITE        304    306       CELL ATTACHMENT SITE.
FT   CONFLICT     60     60       P -> R (IN REF. 4).
FT   CONFLICT    320    320       R -> C (IN REF. 3).
FT   CONFLICT    323    323       H -> D (IN REF. 4).
SQ   SEQUENCE   328 AA;  35137 MW;  C3246328 CRC32;


  Query Match             5.6%;  Score 80;  DB 1;  Length 328;
  Best Local Similarity   27.8%;  Pred. No. 7.3;
  Matches   40;  Conservative   10;  Mismatches   48;  Indels   46;  Gaps    8;

Qy      22 RDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV---------SKMAAAWPSGPS 72
           | |    |:  |   | |||  | |||  | | | :  |          ::||  |  |
Db       4 RVGCPALPLPPP-PLLPLLPLLLLLLGASGGGGGARAEVLFRCPPCTPERLAACGPP-PV 61

Qy      73 APEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKIN----DA 128
           || |                    |||  |||   : | :|       ::|        |
Db      62 APPA------------------AVAAVAGGAR-MPCAEL------VREPGCGCCSVCA 94

Qy     129 TQEPVNCTNYTAH----VSCFPAP 148
             |   |  ||      : |:| |
Db      95 RLEGEACGVYTPRCGQGLRCYPHP 118


RESULT  14
LIPP_PIG
ID   LIPP_PIG       STANDARD;      PRT;   449 AA.
AC   P00591;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TRIACYLGLYCEROL LIPASE, PANCREATIC (EC 3.1.1.3) (PANCREATIC LIPASE)
DE   (PL).
GN   PNLIP.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE OF 308-449.
RX   MEDLINE; 82113655.
RA   DE CARO J.D., BOUDOUARD M., BONICEL J.J., GUIDONI A.A., DESNUELLE P.,
RA   ROVERY M.;
RT   "Porcine pancreatic lipase. Completion of the primary structure.";
RL   Biochim. Biophys. Acta 671:129-138(1981).
RN   [2]
RP   SEQUENCE OF 1-234, AND CARBOHYDRATE-BINDING SITE.
RX   MEDLINE; 79236335.
RA   BIANCHETTA J.D., BIDAUD J., GUIDONI A.A., BONICEL J.J., ROVERY M.;
RT   "Porcine pancreatic lipase. Sequence of the first 234 amino acids of
RT   the peptide chain.";
RL   Eur. J. Biochem. 97:395-405(1979).
RN   [3]
RP   SEQUENCE OF 235-307.
RX   MEDLINE; 80088446.
RA   GUIDONI A.A., BONICEL J.J., BIANCHETTA J.D., ROVERY M.;
RT   "Porcine pancreatic lipase. Sequence between the 235th and 307th
RT   amino acids.";
RL   Biochimie 61:841-845(1979).
RN   [4]
RP   DISULFIDE BONDS.
RX   MEDLINE; 83105095.
RA   BENKOUKA F., GUIDONI A.A., DE CARO J.D., BONICEL J.J.,
RA   DESNUELLE P.A., ROVERY M.;
RT   "Porcine pancreatic lipase. The disulfide bridges and the sulfhydryl
RT   groups.";
RL   Eur. J. Biochem. 128:331-341(1982).
RN   [5]
RP   SUBSTRATE-BINDING SITE.
RX   MEDLINE; 82000578.
RA   GUIDONI A.A., BENKOUKA F., DE CARO J.D., ROVERY M.;
RT   "Characterization of the serine reacting with diethyl p-nitrophenyl
RT   phosphate in porcine pancreatic lipase.";
RL   Biochim. Biophys. Acta 660:148-150(1981).
RN   [6]
RP   STRUCTURE OF CARBOHYDRATE.
RX   MEDLINE; 88082841.
RA   FOURNET B., LEROY Y., MONTREUIL J., DECARO J., ROVERY M.,
RA   VAN KUIK J.A., VLIEGENTHART J.F.G.;
RT   "Primary structure of the glycans of porcine pancreatic lipase.";
RL   Eur. J. Biochem. 170:369-371(1987).
CC   -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC       A FATTY ACID ANION.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
CC       HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/L/PL.html".
DR   PIR; A00732; LIPG.
DR   HSSP; P29183; 1HPL.
DR   PROSITE; PS00120; LIPASE_SER; 1.
DR   PFAM; PF00151; lipase; 1.
DR   PFAM; PF01477; PLAT; 1.
KW   Hydrolase; Lipid degradation; Pancreas; Glycoprotein.
FT   ACT_SITE    152    152       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    176    176       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    263    263       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID      4     10
FT   DISULFID     90    101       IN ISOMER 1.
FT   DISULFID     90    103       IN ISOMER 2.
FT   DISULFID    237    261
FT   DISULFID    285    296
FT   DISULFID    299    304
FT   DISULFID    433    449
FT   CARBOHYD    166    166
SQ   SEQUENCE   449 AA;  49926 MW;  662777B9 CRC32;


  Query Match             5.6%;  Score 80;  DB 1;  Length 449;
  Best Local Similarity   20.2%;  Pred. No. 10;
  Matches   47;  Conservative   30;  Mismatches   96;  Indels   60;  Gaps   10;

Qy       1 MHILKGSPNVIP------------RAHGQKNTRRDG-----TGLYPMRGPFKNLALL--- 41
             : :|| |    |              | |:   | :|     ||| |    |:    |
Db     132 VEVLKSSLGYSPSNVHVIGHSLGSHAAGEAGRRTNGTIERITGLDPAEPCFQGTPELVRL 191

Qy      41 ------------PFSLPLLGGGGSGSGEKVSVSKMAAAWPSG----PSAPEAVTARLVGV 84
                        : |::   | |  : |       :|:|     |   : : :::| :
Db     192 DPSDAKFVDVIHTDAAPIIPNLGFGMSQTVGHLDF---FPNGGKQMPGCQKNILSQIVDI 248

Qy      85 LWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCTNYTAHVS- 144
                 | |       |  : : |  |:  :|   |  :|        | :|    :
Db     249 -------DGIW------EGTRDFVACNHLRSYKYYA-DSILNPDGFAGFPCDSYNVFTAN 294

Qy     144 -CFPAPNITCKDSSGNETHFTG-----NEVGFFKPISCRNVNGYSYKVAVALS 190
            ||| |:  |         | |      ::| :        |   : |||:| ||
Db     295 KCFPCPSEGCPQMGHYADRFPGKTNGVSQVFYLNTGDASNFARWRYKVSVTLS 347


RESULT  15
AUXI_BOVIN
ID   AUXI_BOVIN     STANDARD;      PRT;   910 AA.
AC   Q27974;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   AUXILIN.
```

```
Qy     157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
           || |    |::|        |:| | | :|| |  :|| |     ||      |    |
Db     286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL------LAASLLC 333

Qy     210 LLKFC 214
           ||  |
Db     334 LLSKC 338


RESULT  12
VG32_BPMD2
ID   VG32_BPMD2     STANDARD;      PRT;   223 AA.
AC   O64226;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENE 32 PROTEIN (GP32).
[redacted]32.
[redacted]Mycobacteriophage D29.
OC   Viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98300335.
RA   FORD M.E., SARKIS G.J., BELANGER A.E., HENDRIX R.W., HATFULL G.F.;
RT   "Genome structure of mycobacteriophage D29: implications for phage
RT   evolution.";
RL   J. Mol. Biol. 279:143-164(1998).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF022214; AAC18473.1; -.
SQ   SEQUENCE   223 AA;  21822 MW;  6623D1E6 CRC32;


  Query Match             5.6%;  Score 80.5;  DB 1;  Length 223;
  Best Local Similarity   30.7%;  Pred. No. 4.3;
  Matches   27;  Conservative    8;  Mismatches   36;  Indels   17;  Gaps    3;

Qy      29 PMRGPFKNLALLPFSLP---------LLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTA 79
           |:  |   :    .:::|         ||| || | |        | ||  |      |||
[redacted] 37 PVLTPVTAVGAYTYNIPAQAEFIDVILLGAGGGGQG-----MGSATAWGQGGFGGSWVTA 91

Qy      80 RL---VGVLWFVSVTTGPWGAVATSAGG 104
           |   | : | |:  ||  ||  |:  |
Db      92 TLRRGVDIPWAVTQITGVIGAGGTAGPG 119


RESULT  13
IBP2_HUMAN
ID   IBP2_HUMAN     STANDARD;      PRT;   328 AA.
AC   P18065; Q14619;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 2 PRECURSOR (IGFBP-2)
DE   (IBP-2) (IGF-BINDING PROTEIN 2).
GN   IGFBP2 OR BP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 91293227.
RA   AGARWAL N., HSIEH C.L., SILLS D., SWAROOP M., DESAI B., FRANCKE U.,
RA   SWAROOP A.;
RT   "Sequence analysis, expression and chromosomal localization of a
RT   gene, isolated from a subtracted human retina cDNA library, that
RT   encodes an insulin-like growth factor binding protein (IGFBP2).";
RL   Exp. Eye Res. 52:549-561(1991).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 40-77.
RX   MEDLINE; 90368661.
RA   ZAPF J., KIEFER M., MERRYWEATHER J., MASIARZ F., BAUER D., BORN W.,
RA   FISCHER J.A., FORESCH E.R.;
RT   "Isolation from adult human serum of four insulin-like growth factor
RT   (IGF) binding proteins and molecular cloning of one of them that is
RT   increased by IGF I administration and in extrapancreatic tumor
RT   hypoglycemia.";
RL   J. Biol. Chem. 265:14892-14898(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL LIVER;
RX   MEDLINE; 90060007.
RA   BINKERT C., LANDWEHR J., MARY J.L., SCHWANDER J., HEINRICH G.;
RT   "Cloning, sequence analysis and expression of a cDNA encoding a novel
RT   insulin-like growth factor binding protein (IGFBP-2).";
RL   EMBO J. 8:2497-2502(1989).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 91248211.
RA   EHRENBORG E., VILHELMSDOTTER S., BAJALICA S., LARSSON C., STERM I.,
RA   KOCH J., BRONDUM-NIELSEN K., LUTHMAN H.;
RT   "Structure and localization of the human insulin-like growth factor-
RT   binding protein 2 gene.";
RL   Biochem. Biophys. Res. Commun. 176:1250-1255(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 92293159.
RA   BINKERT C., MARGOT J.B., LANDWEHR J., HEINRICH G., SCHWANDER J.;
RT   "Structure of the human insulin-like growth factor binding protein-2
RT   gene.";
RL   Mol. Endocrinol. 6:826-836(1992).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S37730; AAB22308.1; -.
DR   EMBL; S37712; AAB22308.1; JOINED.
DR   EMBL; S37722; AAB22308.1; JOINED.
DR   EMBL; S37726; AAB22308.1; JOINED.
DR   EMBL; M35410; AAA03246.1; -.
DR   EMBL; X16302; CAA34373.1; -.
DR   EMBL; M69241; AAA36048.1; -.
DR   EMBL; M69237; AAA36048.1; JOINED.
DR   EMBL; M69239; AAA36048.1; JOINED.
DR   EMBL; M69240; AAA36048.1; JOINED.
DR   EMBL; A09809; CAA00862.1; -.
DR   PIR; S05389; S05389.
DR   PIR; A35803; A35803.
DR   PIR; JQ1029; JQ1029.
DR   PIR; A41927; A41927.
DR   PIR; A53957; A53957.
```

```
DR   PIR; A00934; BBHU.
DR   PIR; S14339; S14339.
DR   PIR; S34075; S34075.
DR   HSSP; P20231; 1AAO.
DR   SWISS-2DPAGE; P00751; HUMAN.
DR   MIM; 138470; -.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00084; sushi; 3.
DR   PFAM; PF00089; trypsin; 1.
DR   PFAM; PF00092; vwa; 1.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; MHC III; Repeat; Sushi; Signal; Polymorphism; Zymogen.
FT   SIGNAL       1    25
FT   CHAIN       26   764       COMPLEMENT FACTOR B.
FT   CHAIN       26   259       BA FRAGMENT.
FT   CHAIN      260   764       BB FRAGMENT.
FT   DOMAIN      36   219       3 X SUSHI (SCR) REPEATS.
FT   REPEAT      36    99       SUSHI 1.
FT   REPEAT     102   159       SUSHI 2.
FT   REPEAT     164   219       SUSHI 3.
FT   DOMAIN     482   764       CATALYTIC.
FT   ACT_SITE   526   526       CHARGE RELAY SYSTEM.
FT   ACT_SITE   576   576       CHARGE RELAY SYSTEM.
FT   ACT_SITE   699   699       CHARGE RELAY SYSTEM.
FT   DISULFID    37    76       BY SIMILARITY.
FT   DISULFID    62    98       BY SIMILARITY.
FT   DISULFID   103   145       BY SIMILARITY.
FT   DISULFID   131   158       BY SIMILARITY.
FT   DISULFID   165   205       BY SIMILARITY.
FT   DISULFID   191   218       BY SIMILARITY.
FT   CARBOHYD   122   122
FT   CARBOHYD   142   142
FT   CARBOHYD   285   285
FT   CARBOHYD   378   378
FT   CARBOHYD   291   291       GLUCOSE (GLYCATION).
FT   VARIANT     28    28       W -> R (IN S ALLELE).
FT                              /FTId=VAR_006492.
FT   VARIANT     28    28       W -> E (IN FA ALLELE).
FT                              /FTId=VAR_006493.
FT   VARIANT     32    32       R -> Q (IN S ALLELE).
FT                              /FTId=VAR_006494.
FT   VARIANT    736   736       A -> S (IN FA ALLELE).
FT                              /FTId=VAR_006495.
FT   CONFLICT   297   297       I -> T (IN REF. 7).
FT   CONFLICT   300   300       V -> L (IN CAB23472).
FT   CONFLICT   328   328       D -> V (IN CAB23472).
FT   CONFLICT   356   357       KK -> EE (IN CAB23472).
FT   CONFLICT   537   537       I -> T (IN REF. 9).
FT   CONFLICT   764   764       L -> H (IN REF. 9).
SQ   SEQUENCE   764 AA;  85532 MW;  711DE8BD CRC32;
```

```
  Query Match             5.8%;  Score 83;  DB 1;  Length 764;
  Best Local Similarity   24.1%;  Pred. No. 11;
  Matches   49;  Conservative   21;  Mismatches   71;  Indels   62;  Gaps   12;

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----------GSGSGEKVSV---------- 61
           |: | |       | |:|| | || ||            || | | | :
Db       2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLARPQGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
             :   |||  |   :  | |          :|| |  : |             :: |
Db      56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy     111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
            | : |:|  : |  | : :   :|   ||   |     | ||: :     :|    |
Db     106 PHDFENGEYWPRSPYYNVSDEISFHCVDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy     167 VGFFK----PISCRNVNGYSYKV 185
            |:      ||  | | |  |::
Db     162 AGYCSNPGIPIGTRKV-GSQYRL 183
```

```
RESULT  11
LAMP_HUMAN
ID   LAMP_HUMAN     STANDARD;      PRT;   338 AA.
AC   Q13449;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   LIMBIC SYSTEM-ASSOCIATED MEMBRANE PROTEIN PRECURSOR (LSAMP).
GN   LAMP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96235133.
RA   PIMENTA A.F., FISCHER I., LEVITT P.;
RT   "cDNA cloning and structural analysis of the human limbic-system-
RT   associated membrane protein (LAMP).";
RL   Gene 170:189-195(1996).
CC   -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC       CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC       MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC       OF THE HYPPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON LIMBIC NEURONS AND FIBER TRACTS
CC       AS WELL AS IN SINGLE LAYERS OF THE SUPERIOR COLLICULUS, SPINAL
CC       CHORD AND CEREBELLUM.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE OBCAM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41901; AAC50569.1; -.
DR   MIM; 603241; -.
DR   PFAM; PF00047; ig; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Signal.
FT   SIGNAL       1    28       POTENTIAL.
FT   CHAIN       29   315       LIMBIC SYSTEM-ASSOCIATED MEMBRANE
FT                              PROTEIN.
FT   PROPEP     316   338       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      46   118       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     146   204       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     232   297       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    53   111       POTENTIAL.
FT   DISULFID   153   197       POTENTIAL.
FT   DISULFID   239   290       POTENTIAL.
FT   CARBOHYD    40    40       POTENTIAL.
FT   CARBOHYD    66    66       POTENTIAL.
FT   CARBOHYD   136   136       POTENTIAL.
FT   CARBOHYD   148   148       POTENTIAL.
FT   CARBOHYD   279   279       POTENTIAL.
FT   CARBOHYD   287   287       POTENTIAL.
FT   CARBOHYD   300   300       POTENTIAL.
FT   CARBOHYD   315   315       POTENTIAL.
FT   LIPID      315   315       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   338 AA;  37308 MW;  DF4014E4 CRC32;
```

```
  Query Match             5.7%;  Score 82.5;  DB 1;  Length 338;
  Best Local Similarity   29.6%;  Pred. No. 4.6;
  Matches   37;  Conservative   14;  Mismatches   47;  Indels   27;  Gaps    7;

Qy     101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db     230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285
```

```
          :   |:  :    :|  ::|:   ::|  || |  | : : ||:|     :|   : |:
Db     272 PRDAQCK--SDVFNQVVKSISIVGSYVGNRADTREALDFFSRGLVKAPIKILGLSELASV 329

Qy     225 IDFILISMQIVG 236
           |  ::  ||||
Db     330 YD-KMVKGQIVG 340


RESULT  10
CFAB_HUMAN
ID   CFAB_HUMAN     STANDARD;      PRT;   764 AA.
AC   P00751; O15006; Q29944;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   COMPLEMENT FACTOR B PRECURSOR (EC 3.4.21.47) (C3/C5 CONVERTASE)
DE   (PROPERDIN FACTOR B) (GLYCINE-RICH BETA GLYCOPROTEIN) (GBG) (PBF2).
     BF.
     Homo sapiens (Human).
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A. (ALLELES S; FA AND FB).
RX   MEDLINE; 91065702.
RA   DAVRINCHE C., ABBAL M., CLERC A.;
RT   "Molecular characterization of human complement factor B subtypes.";
RL   Immunogenetics 32:309-312(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ALLELE S).
RC   TISSUE=LIVER;
RX   MEDLINE; 94237735.
RA   MEJIA J.E., JAHN I., DE LA SALLE H., HAUPTMANN G.;
RT   "Human factor B. Complete cDNA sequence of the BF*S allele.";
RL   Hum. Immunol. 39:49-53(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ALLELE S).
RC   TISSUE=LIVER;
RX   MEDLINE; 94041399.
RA   SCHWAEBLE W., LUETTIG B., SOKOLOWSKI T., ESTALLER C., WEISS E.H.,
RA   MEYER ZUM BUESCHENFELDE K.-H., WHALEY K., DIPPOLD W.;
RT   "Human complement factor B: functional properties of a recombinant
RT   zymogen of the alternative activation pathway convertase.";
RL   Immunobiology 188:221-232(1993).
RN   [4]
RP   SEQUENCE FROM N.A. (ALLELE S).
RX   MEDLINE; 94067177.
     HORIUCHI T., KIM S., MATSUMOTO M., WATANABE I., FUJITA S.,
     VOLANAKIS J.E.;
RT   "Human complement factor B: cDNA cloning, nucleotide sequencing,
RT   phenotypic conversion by site-directed mutagenesis and expression.";
RL   Mol. Immunol. 30:1587-1592(1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   ROWEN L., DANKERS C., BASKIN D., FAUST J., LORETZ C., AHEARN M.E.,
RA   BANTA A., SWARTZELL S., SMITH T.M., SPIES T., HOOD L.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 26-764, PARTIAL SEQUENCE FROM N.A., AND CARBOHYDRATES.
RX   MEDLINE; 84161997.
RA   MOLE J.E., ANDERSON J.K., DAVISON E.A., WOODS D.E.;
RT   "Complete primary structure for the zymogen of human complement
RT   factor B.";
RL   J. Biol. Chem. 259:3407-3412(1984).
RN   [7]
RP   SEQUENCE OF 260-764.
RX   MEDLINE; 83204002.
RA   CHRISTIE D.L., GAGNON J.;
RT   "Amino acid sequence of the Bb fragment from complement Factor B.
RT   Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and
RT   completion of the sequence of the Bb fragment.";
RL   Biochem. J. 209:61-70(1983).
RN   [8]
RP   SEQUENCE OF 339-764 FROM N.A.
RX   MEDLINE; 83273641.
RA   CAMPBELL R.D., PORTER R.R.;
RT   "Molecular cloning and characterization of the gene coding for human
RT   complement protein factor B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:4464-4468(1983).
RN   [9]
RP   SEQUENCE OF 467-595 AND 752-764 FROM N.A.
RX   MEDLINE; 83039428.
RA   WOODS D.E., MARKHAM A.F., RICKER A.T., GOLDBERGER G., COLTEN H.R.;
RT   "Isolation of cDNA clones for the human complement protein factor B,
RT   a class III major histocompatibility complex gene product.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:5661-5665(1982).
RN   [10]
RP   SEQUENCE OF 16-259 FROM N.A.
RX   MEDLINE; 84158524.
RA   MORLEY B.J., CAMPBELL R.D.;
RT   "Internal homologies of the Ba fragment from human complement
RT   component Factor B, a class III MHC antigen.";
RL   EMBO J. 3:153-157(1984).
RN   [11]
RP   SEQUENCE OF 1-99 FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 87102880.
RA   WU L.C., MORLEY B.J., CAMPBELL R.D.;
RT   "Cell-specific expression of the human complement protein factor B
RT   gene: evidence for the role of two distinct 5'-flanking elements.";
RL   Cell 48:331-342(1987).
RN   [12]
RP   GLYCATION IN POSITION 291.
RX   MEDLINE; 91174758.
RA   NIEMANN M.A., BHOWN A.S., MILLER E.J.;
RT   "The principal site of glycation of human complement factor B.";
RL   Biochem. J. 274:473-480(1991).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE. IT HAS ALSO BEEN IMPLICATED
CC       IN PROLIFERATION AND DIFFERENTIATION OF PREACTIVATED B
CC       LYMPHOCYTES, RAPID SPREADING OF PERIPHERAL BLOOD MONOCYTES,
CC       STIMULATION OF LYMPHOCYTE BLASTOGENESIS AND LYSIS OF ERYTHROCYTES.
CC       BA INHIBITS THE PROLIFERATION OF PREACTIVATED B LYMPHOCYTES.
CC   -!- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND
CC       C3B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH
CC       CLEAVAGES TAKE PLACE AT THE C-TERMINAL OF AN ARGININE RESIDUE.
CC   -!- SUBUNIT: MONOMER.
CC   -!- POLYMORPHISM: TWO MAJOR VARIANTS, F AND S, AND 2 MINOR VARIANTS,
CC       AS WELL AS AT LEAST 14 VERY RARE VARIANTS, HAVE BEEN IDENTIFIED.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS-III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) REPEATS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72875; CAA51389.1; -.
DR   EMBL; S67310; AAD13989.1; -.
DR   EMBL; L15702; AAA16820.1; -.
DR   EMBL; X00284; CAA25077.1; -.
DR   EMBL; AF019413; AAB67977.1; -.
DR   EMBL; K01566; CAB23472.1; -.
DR   EMBL; J00125; -; NOT_ANNOTATED_CDS.
DR   EMBL; J00126; AAA36226.1; -.
DR   EMBL; J00185; AAA36219.1; ALT_SEQ.
DR   EMBL; J00186; AAA36220.1; -.
DR   EMBL; M15082; AAA59625.1; -.
```

```
Qy     210 LLKFCTVGFC 219
             || :  |
Db     329 ASLFCLLSKC 338


RESULT   8
Y346_MYCTU
ID   Y346_MYCTU     STANDARD;      PRT;   487 AA.
AC   O06297;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 52.2 KD TRANSPORT PROTEIN RV0346C.
GN   RV0346C OR MTCY13E10.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
     STRAIN=H37RV;
     MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95324; CAB08578.1; -.
DR   PROSITE; PS00218; AMINO_ACID_PERMEASE; 1.
DR   PFAM; PF00324; aa_permeases; 1.
KW   Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
FT   TRANSMEM     26     46       POTENTIAL.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     98    118       POTENTIAL.
FT   TRANSMEM    133    153       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    256    276       POTENTIAL.
FT   TRANSMEM    290    310       POTENTIAL.
FT   TRANSMEM    341    361       POTENTIAL.
FT   TRANSMEM    369    389       POTENTIAL.
FT   TRANSMEM    414    434       POTENTIAL.
FT   TRANSMEM    440    460       POTENTIAL.
SQ   SEQUENCE   487 AA;  52194 MW;  64BBBCCD CRC32;


  Query Match             5.8%;  Score 83.5;  DB 1;  Length 487;
  Best Local Similarity   26.5%;  Pred. No. 5.7;
  Matches   27;  Conservative   16;  Mismatches   40;  Indels   19;  Gaps    5;

Qy     159 ETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGF 218
           :   |  : |: | :  | :   :   |:  ||||     |     | |    |   |
Db       8 DERLTREDTGYHKGLHSRQLQMIALGGAIGTGLFLG--AGGRLASAGPGL----FLVYGI 61

Qy     219 CGIGSLIDFILISMQIVG------PSDGS--SYIIDYYGTRL 252
                ||| |: : :: :|      || ||  ||  ::|| ::
Db      62 CGI-----FVFLILRALGELVLHRPSSGSFVSYAREFYGEKV 98


RESULT   9
ADH1_KLUMA
ID   ADH1_KLUMA     STANDARD;      PRT;   348 AA.
AC   Q07288;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   ALCOHOL DEHYDROGENASE 1 (EC 1.1.1.1).
GN   ADH1.
OS   Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Kluyveromyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12424;
RX   MEDLINE; 93250057.
RA   LADRIERE J.M., DELCOUR J., VANDENHAUTE J.;
RT   "Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase
RT   from Kluyveromyces marxianus ATCC 12424.";
RL   Biochim. Biophys. Acta 1173:99-101(1993).
CC   -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60224; CAA42785.1; -.
DR   PIR; S32521; S32521.
DR   PROSITE; PS00059; ADH_ZINC; 1.
DR   PFAM; PF00107; adh_zinc; 1.
KW   Oxidoreductase; Zinc; NAD; Multigene family.
FT   METAL        44     44       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        67     67       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        98     98       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       101    101       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       104    104       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       112    112       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       154    154       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   348 AA;  37158 MW;  FA9C656F CRC32;


  Query Match             5.8%;  Score 83;  DB 1;  Length 348;
  Best Local Similarity   20.8%;  Pred. No. 4.3;
  Matches   65;  Conservative   38;  Mismatches  107;  Indels  102;  Gaps   14;

Qy      19 NTRRDG---TGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSG----- 71
           | :  |   | |:  :| :     |   |||: ||  |:|  |:: :    |  |
Db      37 NVKYSGVCHTDLHAWQGDWP----LDTKLPLV-GGHEGAGIVVAMGENVTGWEIGDYAGI 91

Qy      71 -----------------PSAPEA-------------VTARLV---------------- 83
                            |: |:|              ||  |
Db      92 KWLNGSCMSCEECELSNEPNCPKADLSGYTHDGSFQQYATADAVQAARIPKNVDLAEVAP 151

Qy      83 ----GV-----LWFVSVTTGPWGAVATSAGGEESLKCEDLKV----------------- 116
               ||     |    :  | | |:: : ||  ||  :  |
Db     152 ILCAGVTVYKALKSAHIKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDAGDEKAKLF 211

Qy     116 ----GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE--VGF 169
               |:|     |  |   | :  ||  ||     : :     :|   |   |    ||
Db     212 KELGGEYFIDFTKTKDMVAEVIEATNGVAHAVINVSVSEAAISTSVLYTRSNGTVVLVGL 271

Qy     170 FKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLK--FCTVGFCGIGSL 224
```

```
Db    200 DAFYFHFYQHGGTFYAYYADKPSATTFLFSVY----IGDILTQYYVLPFICNPTAGSTFA 255

Qy    245 IDYYGTRLTR 254
          |: | | :
Db    256 PRYWVTPLVK 265


RESULT   4
EF1S_PORPU
ID   EF1S_PORPU     STANDARD;      PRT;   515 AA.
AC   P50257;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ELONGATION FACTOR 1-ALPHA S (EF-1-ALPHA S) (SPOROPHYTE-SPECIFIC EF-1-
DE   ALPHA).
GN   TEF-S.
OS   Porphyra purpurea.
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AVONPORT;
RX   MEDLINE; 96309386.
RA   LIU Q.Y., BALDAUF S.L., REITH M.E.;
RT   "Elongation factor 1 alpha genes of the red alga Porphyra purpurea
RT   include a novel, developmentally specialized variant.";
RL   Plant Mol. Biol. 31:77-85(1996).
CC   -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC       AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC       BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY IN THE SPOROPHYTE, A SHELL-
CC       BORING, FILAMENTOUS PHASE.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-TU/EF-1A SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08841; AAA61790.1; -.
DR   HSSP; Q01698; 1TUI.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
DR   PFAM; PF00009; GTP_EFTU; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Multigene family.
FT   NP_BIND      14     21       GTP (BY SIMILARITY).
FT   NP_BIND      91     95       GTP (BY SIMILARITY).
FT   NP_BIND     151    154       GTP (BY SIMILARITY).
SQ   SEQUENCE   515 AA;  56648 MW;  03F72E83 CRC32;


  Query Match             6.2%;  Score 89.5;  DB 1;  Length 515;
  Best Local Similarity   24.7%;  Pred. No. 1.9;
  Matches   37;  Conservative   20;  Mismatches   56;  Indels   37;  Gaps    6;

Qy     41 PFSLPL-----LGGGGSGSGEKVSVSKMAAAW-----PSGPSAPEAVTARLVGVLWFVSV 90
          |  |||     :|| |:    :|    : |       |:| :| |  :  :
Db    261 PLRLPLQDVYKIGGIGTVPVGRVETGILKAGMQVTFEPAGKAAVEVKSVEM-------HH 313

Qy     91 TTGPWGAVATSAGGEESLKCEDLKVGQYICKDPK----------INDATQEPVNCTN-- 138
          |: |      : |    |  :|:| |  :| | |           |   |:  |  |
Db    314 TSVPQAIPGDNVGFNVKLTVKDIKRGD-VCGDTKNDPPIPTECFLANVIIQDHKNIRNGY 372

Qy    138 ------YTAHVSCFPAPNITCKDSSGNETH 161
                :|||::|  |  :: ||  | :||
Db    373 TPVLDCHTAHIACKFASILSKKDKRGKQTH 402


RESULT   5
LCT2_MOUSE
ID   LCT2_MOUSE     STANDARD;      PRT;   151 AA.
AC   O88803; O88804;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   LEUKOCYTE CELL-DERIVED CHEMOTAXIN 2 PRECURSOR.
GN   LECT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=LIVER;
RX   MEDLINE; 98382586.
RA   YAMAGOE S., WATANABE T., MIZUNO S., SUZUKI K.;
RT   "The mouse Lect2 gene: cloning of cDNA and genomic DNA, structural
RT   characterization and chromosomal localization.";
RL   Gene 216:171-178(1998).
CC   -!- FUNCTION: HAS A NEUTROPHIL CHEMOTACTIC ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE LECT2 / MIM-1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB009687; BAA33383.1; -.
DR   EMBL; AB009688; BAA33384.1; -.
DR   MGD; MGI:1278342; LECT2.
KW   Chemotaxis; Signal; Alternative splicing.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    151       LEUKOCYTE CELL-DERIVED CHEMOTAXIN 2.
FT   VARSPLIC     98    151       FCVKIFYIKPIKYKGSIKKGEKLGTLLPLQKIYPGIQSHVH
FT                                VENCDSSDPTAYL -> QRLQAHTTTLNVFTCYWDKIQIPR
FT                                PTRFLCQNFLH (IN ISOFORM LECT2Q).
SQ   SEQUENCE   151 AA;  16405 MW;  C5BD189A CRC32;


  Query Match             6.0%;  Score 86.5;  DB 1;  Length 151;
  Best Local Similarity   24.8%;  Pred. No. 0.83;
  Matches   29;  Conservative    9;  Mismatches   52;  Indels   27;  Gaps    4;

Qy     78 TARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCTN 137
          |  |:       |   |||  :  |    |   |:    |||  :       ||
Db      4 TTILISAALLSSALAGPWANICASKSSNEIRTCDSYGCGQYSAQ------RTQR------ 52

Qy    138 YTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNV---------NGYSYKV 185
               |    |  :: | | |     |||  ||  ||   :|           |:  |:
Db     52 ---H---HPGVDVLCSDGSVVYAPFTGKIVGQEKPYRNKNAINDGIRLSGRGFCVKI 102


RESULT   6
ADH2_KLULA
ID   ADH2_KLULA     STANDARD;      PRT;   348 AA.
AC   P49383;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ALCOHOL DEHYDROGENASE II (EC 1.1.1.1).
GN   ADH2.
OS   Kluyveromyces lactis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Kluyveromyces.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
Db    466 MGGYRAEDVEFTN----VKCRVLPCIEC---HGPRT--------FTKSTPCIIYNGHYFL 510

Qy    185 VAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSY 243
          :  |:||| :  ||| ||| |: : |  |:|  ||  ::|  |: : ::||:| ||:
Db    511 TTLLYSIFLGVVAVDRFCLGYSAMAVGKLMTLGGFGIWWIVDIFLLVLGVLGPADDSSW 569


RESULT   2
YJ05_CAEEL
ID   YJ05_CAEEL     STANDARD;      PRT;   753 AA.
AC   O01975;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 85.1 KD NUCLEASE C41D11.5 (EC 3.1.30.-).
GN   C41D11.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GATTUNG S., MAGGI L.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE DNA/RNA NON-SPECIFIC ENDONUCLEASES
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF003740; AAC48141.1; -.
DR   WORMPEP; C41D11.5; CE08662.
DR   PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR   PFAM; PF01223; Endonuclease; 1.
KW   Hypothetical protein; Hydrolase; Nuclease; Endonuclease.
FT   ACT_SITE    593    593       BY SIMILARITY.
SQ   SEQUENCE   753 AA;  85068 MW;  9DF3803A CRC32;


  Query Match             11.1%;  Score 159.5;  DB 1;  Length 753;
  Best Local Similarity   28.2%;  Pred. No. 2.7e-06;
  Matches   46;  Conservative   27;  Mismatches   61;  Indels   29;  Gaps    5;

Qy    104 GEESLKCE---DLKVGQYICKDPKINDATQE----PVNCTNYTA----------HVSCF 145
          |   | |    | ::|  :    |:|  :::    ||:  |  |           |
Db    284 GSAGLTCTFPGDCRIGDTV----KVNCTSRKGCPNPVSRNNVEAVCRFCWQLLPGDYDCE 339

Qy    146 PAPNITCKDS-------SGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGA 198
          || | :   :       | : :     :  |:| | |   :|||:   : ||: ||  ||
Db    340 PATNCSTSSTKLLVTKCSAHSSVICMGQRNFYKRIPCNWSSGYSWTKTMILSVVLGGFGA 399

Qy    199 DRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGS 241
          ||||||     : |  : |  |: :|:| :||::  : | |||
Db    400 DRFYLGLWKSAIGKLFSFGGLGVWTLVDVVLIAVGYIKPYDGS 442


RESULT   3
VGL2_CVMA5
ID   VGL2_CVMA5     STANDARD;      PRT;  1324 AA.
AC   P11224;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE   [CONTAINS: SPIKE PROTEIN S1 (90B); SPIKE PROTEIN S2 (90A)].
GN   S.
OS   Murine coronavirus MHV (strain A59).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88072088.
RA   LUYTJES W., STURMAN L.S., BREDENBEEK P.J., CHARITE J.,
RA   VAN DER ZEIJST B.A.M., HORZINEK M.C., SPAAN W.J.M.;
RT   "Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and
RT   identification of the trypsin cleavage site.";
RL   Virology 161:479-487(1987).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC       AND IN SYNCYTIUM FORMATION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M18379; AAA46455.1; -.
DR   PIR; A27402; VGIH59.
DR   PFAM; PF01601; Corona_S2; 2.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17   1324       E2 GLYCOPROTEIN.
FT   CHAIN        17    717       SPIKE PROTEIN S1.
FT   CHAIN       718   1324       SPIKE PROTEIN S2.
FT   DOMAIN       17   1265       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1266   1286       POTENTIAL.
FT   DOMAIN     1287   1324       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1287   1304       CYS-RICH.
FT   CARBOHYD     31     31       POTENTIAL.
FT   CARBOHYD     60     60       POTENTIAL.
FT   CARBOHYD    192    192       POTENTIAL.
FT   CARBOHYD    357    357       POTENTIAL.
FT   CARBOHYD    435    435       POTENTIAL.
FT   CARBOHYD    530    530       POTENTIAL.
FT   CARBOHYD    625    625       POTENTIAL.
FT   CARBOHYD    657    657       POTENTIAL.
FT   CARBOHYD    665    665       POTENTIAL.
FT   CARBOHYD    688    688       POTENTIAL.
FT   CARBOHYD    737    737       POTENTIAL.
FT   CARBOHYD    754    754       POTENTIAL.
FT   CARBOHYD    893    893       POTENTIAL.
FT   CARBOHYD   1180   1180       POTENTIAL.
FT   CARBOHYD   1190   1190       POTENTIAL.
FT   CARBOHYD   1209   1209       POTENTIAL.
FT   CARBOHYD   1225   1225       POTENTIAL.
FT   CARBOHYD   1246   1246       POTENTIAL.
SQ   SEQUENCE   1324 AA;  146019 MW;  B23B6829 CRC32;


  Query Match              6.3%;  Score 90;  DB 1;  Length 1324;
  Best Local Similarity   23.6%;  Pred. No. 5;
  Matches   59;  Conservative   27;  Mismatches   94;  Indels   70;  Gaps   13;

Qy     25 TGLYPMRG-PFKNLALLP--------FSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSA-- 74
          || ||: |  |:||||           |  | |  |     |   |  |  :  : ||.| :|
Db     66 TGYYPVDGSKFRNLALRGTNSVSLSWFQPPYLNQFNDGIFAK--VQNLKTSTPSGATAYF 123

Qy     74 PEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQY-ICKDPKINDATQEP 132
          |  |   | |   : :|   |:  |  ::           | || ||: |
Db    124 PTIVIGSLFGYTSY-TVVIEPYNGVIMAS-----------VCQYTICQLP--------- 162

Qy    133 VNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE-VGFF----KPISCRNVNGYSYKVAV 187
                :  | |  |              ||: :||:    ||  |     ::  |
Db    162 --------YTDCKPNTN-------------GNKLIGFWHTDVKPPICVLKRNFTLNVNA 199

Qy    188 ALSLFLGWLGADRFYLGY---PALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
                |  :     ||  |   |:     |       || ::    :   |  |: ||::
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 04:22:47 ; Search time 10.6 Seconds
(without alignments)
757.891 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 1439
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 82229 seqs, 29864866 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 82229

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	167.5	11.6	573	1	YKK3_CAEEL	P34280 caenorhabdi
2	159.5	11.1	753	1	YJ05_CAEEL	O01975 caenorhabdi
3	90	6.3	1324	1	VGL2_CVMA5	P11224 murine coro
4	89.5	6.2	515	1	EF1S_PORPU	P50257 porphyra pu
5	86.5	6.0	151	1	LCT2_MOUSE	O88803 mus musculu
6	85.5	5.9	348	1	ADH2_KLULA	P49383 kluyveromyc
7	85	5.9	338	1	LAMP_RAT	Q62813 rattus norv
8	83.5	5.8	487	1	Y346_MYCTU	O06297 mycobacteri
9	83	5.8	348	1	ADH1_KLUMA	Q07288 kluyveromyc
10	83	5.8	764	1	CFAB_HUMAN	P00751 homo sapien
11	82.5	5.7	338	1	LAMP_HUMAN	Q13449 homo sapien
12	80.5	5.6	223	1	VG32_BPMD2	O64226 mycobacteri
13	80	5.6	328	1	IBP2_HUMAN	P18065 homo sapien
14	80	5.6	449	1	LIPP_PIG	P00591 sus scrofa
15	79.5	5.5	910	1	AUXI_BOVIN	Q27974 bos taurus
16	79.5	5.5	4367	1	DYHC_NEUCR	P45443 neurospora
17	79.5	5.5	855	1	POLG_HRV3	Q82081 human rhino
18	79.5	5.5	3432	1	POLG_JAEV1	P27395 j genome po
19	79.5	5.5	3432	1	POLG_JAEV5	P19110 j genome po
20	79.5	5.5	3432	1	POLG_JAEVJ	P32886 j genome po
21	79	5.5	468	1	LIP2_MOUSE	P17892 mus musculu
22	79	5.5	525	1	STP_SPIOL	Q03411 spinacia ol
23	78.5	5.5	245	1	CRS3_HORSE	O19010 equus cabal
24	78.5	5.5	398	1	MUB1_XENLA	P38565 xenopus lae
25	78	5.4	557	1	YP85_MYCTU	Q50636 mycobacteri
26	77.5	5.4	2179	1	POLG_HRV14	P03303 human rhino
27	77	5.4	1239	1	NRG_DROME	P20241 drosophila
28	76.5	5.3	772	1	PMIP_YEAST	P35999 saccharomyc
29	76	5.3	467	1	LIP1_HUMAN	P54315 homo sapien
30	76	5.3	1556	1	VG67_HSV11	Q00107 ictalurid h
31	75.5	5.2	350	1	ADH1_KLULA	P20369 kluyveromyc
32	75.5	5.2	592	1	CEA_CITFR	P04480 citrobacter
33	75.5	5.2	1114	1	DEFI_CHICK	P36197 gallus gall
34	75.5	5.2	682	1	NISP_LACLA	Q07596 lactococcus
35	75.5	5.2	484	1	PER2_VOLCA	P81132 volvox cart
36	75.5	5.2	676	1	PER_DROMA	Q03353 drosophila
37	75.5	5.2	1224	1	PER_DROME	P07663 drosophila
38	75.5	5.2	661	1	PER_DROSE	Q03354 drosophila
39	75.5	5.2	656	1	PER_DROSI	Q03355 drosophila
40	75	5.2	435	1	EF1A_CRYPV	P90519 cryptospori
41	75	5.2	2483	1	PCX_DROME	P18490 drosophila
42	75	5.2	357	1	SFA1_STRFR	Q03424 streptomyce
43	74.5	5.2	375	1	ADH3_YEAST	P07246 saccharomyc
44	74.5	5.2	1256	1	ATL_STAAU	P52081 staphylococ
45	74.5	5.2	596	1	DIVJ_CAUCR	Q03228 caulobacter

ALIGNMENTS

RESULT 1
YKK3_CAEEL
ID YKK3_CAEEL STANDARD; PRT; 573 AA.
AC P34280;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 64.3 KD GTP-BINDING PROTEIN C02F5.3 IN CHROMOSOME III.
GN C02F5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: BELONGS TO THE GTP1/OBG FAMILY.

DR EMBL; L14745; AAA27918.1; -.
DR PIR; S44605; S44605.
DR WORMPEP; C02F5.3; CE00039.
DR PROSITE; PS00905; GTP1_OBG; 1.
DR PFAM; PF01018; GTP1_OBG; 1.
KW Hypothetical protein; GTP-binding.
FT NP_BIND 69 76 GTP (BY SIMILARITY).
FT NP_BIND 115 119 GTP (BY SIMILARITY).
FT NP_BIND 246 249 GTP (BY SIMILARITY).
SQ SEQUENCE 573 AA; 64299 MW; 7B4A73EC CRC32;

Query Match 11.6%; Score 167.5; DB 1; Length 573;
Best Local Similarity 27.9%; Pred. No. 4e-07;
Matches 50; Conservative 23; Mismatches 57; Indels 49; Gaps 5;

```
Qy     90 VTTGPWGAVATSAGGEESLKCEDLKVGQYICKDP----------------------K 124
          |:| | | |         ::|  |:    :|:||                      |
Db    415 VSTNPLGPV---------VECRFLENSFILCEDPVPLYGPGQTGQQPANESFRNEGKCLK 465

Qy    125 INDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYK 184
          :    | |  ||    | |   | | |    | |       | |   |   ||: :
```

submitted to the EMBL Data Library, August 1998
A;Description: The sequence of A. thaliana T24H24.
A;Reference number: Z14333
A;Accession: T01466
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-171 <COU>
A;Cross-references: EMBL:AF075598; NID:g3293581; PIDN:AAC28213.1; PID:g3377840
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 98/1; 116/2
A;Note: T24H24.11
C;Superfamily: Yeast nitrogen fixation protein; nitrogen fixation protein homology
F;61-113/Domain: nitrogen fixation protein homology <NFH>

Query Match 5.6%; Score 80; DB 2; Length 171;
Best Local Similarity 26.5%; Pred. No. 5.9;
Matches 39; Conservative 16; Mismatches 62; Indels 30; Gaps 7;

```
Qy       8 PNVI-----PRAHG--QKNTRRDGTGLY--PMRGPFKNLALLPFSLPLLGGGGS------ 53
           ||||     ||  |   ||    ||||   |  |    | :        | |
Db      30 PNVIDHYDNPRNVGSFDKNDPNVGTGLVGAPQCGDVMKLQV------KFDGSGQIIDAKF 83

Qy      53 ---GSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLK 109
             | |  :: | :|  |  | |  | :| :   :   :|:   |   :  |   |:::|
Db      84 KTFGCGSAIAASSVATEWVKGKSVEEVLTIKNSQIAKHLSL---PPVKLHCSMLAEDAIK 140

Qy     110 CEDLKVGQYICKDPKINDATQEPVNCT 136
               :  |  |  | |  | | :: |
Db     141 A---AIKNYKEKQDKANGETVETIDST 164
```

Search completed: February 11, 2000, 07:38:37
Job time: 878 sec

```
A;Accession: A90675
A;Molecule type: protein
A;Residues: 235-307 <GUI>
A;Experimental source: pancreas
A;Note: this sequence has since been revised at positions 302 and 305
R;Benkouka, F.; Guidoni, A.A.; De Caro, J.D.; Bonicel, J.J.; Desnuelle, P.A.; Rovery, M.
Eur. J. Biochem. 128, 331-341, 1982
A;Title: Porcine pancreatic lipse. The disulfide bridges and the sulfhydryl groups.
A;Reference number: A91124; MUID:83105095
A;Contents: disulfide bonds
A;Accession: A91124
A;Molecule type: protein
A;Residues: 1-182,'E',184-449 <BEN>
A;Experimental source: pancreas
R;Guidoni, A.; Benkouka, F.; De Caro, J.; Rovery, M.
Biochim. Biophys. Acta 660, 148-150, 1981
A;Title: Characterization of the serine reacting with diethyl p-nitrophenyl phosphate in
A;Reference number: A90634; MUID:82000578
A;Contents: annotation; substrate-binding site
A;Comment: Ser-152 reacts with emulsified or micellar diethyl p-nitrophenyl phosphate an
C;Superfamily: triacylglycerol lipase
C;Keywords: carboxylic ester hydrolase; glycoprotein; lipid digestion
F;4-10,237-261,285-296,299-304,433-449/Disulfide bonds: #status experimental
F;90-101/Disulfide bonds: (or 90-103) #status experimental
F;166/Binding site: carbohydrate (Asn) (covalent) #status experimental


  Query Match             5.6%;  Score 80;  DB 1;  Length 449;
  Best Local Similarity   20.2%;  Pred. No. 18;
  Matches   47;  Conservative   30;  Mismatches   96;  Indels   60;  Gaps   10;

Qy       1 MHILKGSPNVIP------------RAHGQKNTRRDG-----TGLYPMRGPFKNLALL--- 41
           : :|| |    |              | |:   | :|     ||| |    |:    |
Db     132 VEVLKSSLGYSPSNVHVIGHSLGSHAAGEAGRRTNGTIERITGLDPAEPCFQGTPELVRL 191

Qy      41 ------------PFSLPLLGGGGSGSGEKVSVSKMAAAWPSG----PSAPEAVTARLVGV 84
                      : |::   | |  : |        :|:|    |   : : :::| :
Db     192 DPSDAKFVDVIHTDAAPIIPNLGFGMSQTVGHLDF---FPNGGKQMPGCQKNILSQIVDI 248

Qy      85 LWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCTNYTAHVS- 144
                | |       |  : : |  |:  :|   |  :|        | :|    :
Db     249 -------DGIW------EGTRDFVACNHLRSYKYYA-DSILNPDGFAGFPCDSYNVFTAN 294

Qy     144 -CFPAPNITCKDSSGNETHFTG-----NEVGFFKPISCRNVNGYSYKVAVALS 190
           ||| |:  |          | |      ::| :        |   : |||:| ||
Db     295 KCFPCPSEGCPQMGHYADRFPGKTNGVSQVFYLNTGDASNFARWRYKVSVTLS 347


RESULT  14
A41927
insulin-like growth factor-binding protein 2 precursor - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C;Accession: A41927; JQ1029; A53957; S05389; A35803; C54651
R;Binkert, C.; Margot, J.B.; Landwehr, J.; Heinrich, G.; Schwander, J.
Mol. Endocrinol. 6, 826-836, 1992
A;Title: Structure of the human insulin-like growth factor binding protein-2 gene.
A;Reference number: A41927; MUID:92293159
A;Accession: A41927
A;Molecule type: DNA
A;Residues: 1-328 <BIN>
A;Cross-references: GB:S37730; NID:g250198; PIDN:AAB22308.1; PID:g250199
A;Experimental source: placenta
A;Note: sequence extracted from NCBI backbone (NCBIN:106610, NCBIN:106613, NCBIN:106615,
R;Ehrenborg, E.; Vilhelmsdotter, S.; Bajalica, S.; Larsson, C.; Sterm, I.; Koch, J.; Bro
Biochem. Biophys. Res. Commun. 176, 1250-1255, 1991
A;Title: Structure and localization of the human insulin-like growth factor-binding prot
A;Reference number: JQ1029; MUID:91248211
A;Accession: JQ1029
A;Molecule type: DNA
A;Residues: 1-59,'R',61-322,'D',324-328 <EHR>
A;Cross-references: GB:M69237
A;Note: the authors translated the codon CGG for residue 60 as Pro and the codon GAC for
R;Agarwal, N.; Hsieh, C.L.; Sills, D.; Swaroop, M.; Desai, B.; Francke, U.; Swaroop,
Exp. Eye Res. 52, 549-561, 1991
A;Title: Sequence analysis, expression and chromosomal localization of a gene, isolat
).
A;Reference number: A53957; MUID:91293227
A;Accession: A53957
A;Molecule type: mRNA
A;Residues: 1-328 <AGA>
A;Cross-references: GB:M35410; NID:g179476; PIDN:AAA03246.1; PID:g179477
A;Experimental source: retina
R;Binkert, C.; Landwehr, J.; Mary, J.L.; Schwander, J.; Heinrich, G.
EMBO J. 8, 2497-2502, 1989
A;Title: Cloning, sequence analysis and expression of a cDNA encoding a novel insulin
A;Reference number: S05389; MUID:90060007
A;Accession: S05389
A;Molecule type: mRNA
A;Residues: 1-319,'C',321-328 <BI2>
A;Cross-references: EMBL:X16302; NID:g33009; PIDN:CAA34373.1; PID:g33010
R;Zapf, J.; Kiefer, M.; Merryweather, J.; Masiarz, F.; Bauer, D.; Born, W.; Fischer,
J. Biol. Chem. 265, 14892-14898, 1990
A;Title: Isolation from adult human serum of four insulin-like growth factor (IGF) bi
ic tumor hypoglycemia.
A;Reference number: A35803; MUID:90368661
A;Accession: A35803
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-19,'P',21-328 <ZAP>
R;Roghani, M.; Segovia, B.; Whitechurch, O.; Binoux, M.
Growth Regul. 1, 125-130, 1991
A;Title: Purification from human cerebrospinal fluid of insulin-like growth factor bi
A;Reference number: A54651; MUID:93091816
A;Accession: C54651
A;Molecule type: protein
A;Residues: 43-44,'X',46-47,'X',49-55,'X',57-61 <ROG>
A;Experimental source: cerebrospinal fluid
A;Note: sequence extracted from NCBI backbone (NCBIP:121072)
C;Genetics:
A;Gene: GDB:IGFBP2; IBP2
A;Cross-references: GDB:126735; OMIM:146731
A;Map position: 2q33-2q34
A;Introns: 151/1; 227/3; 274/3
C;Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I rep
C;Keywords: plasma
F;1-39/Domain: signal sequence #status predicted <SIG>
F;40-328/Product: insulin-like growth factor-binding protein 2 #status predicted <MAT
F;230-309/Domain: thyroglobulin type I repeat homology <THY1>
F;304-306/Region: cell attachment (R-G-D) motif


  Query Match             5.6%;  Score 80;  DB 1;  Length 328;
  Best Local Similarity   27.8%;  Pred. No. 12;
  Matches   40;  Conservative   10;  Mismatches   48;  Indels   46;  Gaps    8;

Qy      22 RDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV---------SKMAAAWPSGPS 72
             | |     |:  |   | |||  | |||  | | | :  |           ::||  |  |
Db       4 RVGCPALPLPPP-PLLPLLPLLLLLGASGGGGGARAEVLFRCPPCTPERLAACGPP-PV 61

Qy      73 APEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKIN----DA 128
           || |                    |||  |||   : | :|        ::|        |
Db      62 APPA-------------------AVAAVAGGAR-MPCAEL------VREPGCGCCSVCA 94

Qy     129 TQEPVNCTNYTAH----VSCFPAP 148
             |   |  ||      : |:| |
Db      95 RLEGEACGVYTPRCGQGLRCYPHP 118


RESULT  15
T01466
nifU protein homolog T24H24.11 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Jul-1999
C;Accession: T01466
R;Courtney, L.; Stoneking, T.; Langston, Y.; Mead, K.
```

A;Reference number: JC4776; MUID:96235133
A;Accession: JC4776
A;Molecule type: mRNA
A;Residues: 1-338 <PIM>
A;Cross-references: GB:U41901; NID:g1276898; PID:g1276899
A;Experimental source: brain
C;Comment: This protein is a neuronal surface glycoprotein distributed in cortical and s
mily. It plays development and function in limbic system.
C;Genetics:
A;Gene: lamp
C;Keywords: brain; glycoprotein; membrane protein; phosphoprotein
F;1-7/Domain: signal sequence #status predicted <SIG>
F;333-338/Region: hydrophobic
F;40,66,136,148,279,287,300,315/Binding site: carbohydrate (Asn) (covalent) #status pred
F;42,115,142,164,171,220,231/Binding site: phosphate (Thr) (covalent) #status predicted
F;95,192,204,236,310/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 5.7%; Score 82.5; DB 2; Length 338;
Best Local Similarity 29.6%; Pred. No. 7.8;
Matches 37; Conservative 14; Mismatches 47; Indels 27; Gaps 7;

```
Qy    101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
          : | : |||||   |     ::   | :|| |    :  |      |     |:| ::
Db    230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy    157 GNETHFTGNEVG-------FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
          || |    |::|       |:| | | :|| |  :|| |     ||      |    |
Db    286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL------LAASLLC 333

Qy    210 LLKFC 214
          ||  |
Db    334 LLSKC 338
```

RESULT 11
T03788
laccase (EC 1.10.3.2) - common tobacco (fragment)
N;Alternate names: diphenol oxidase; urishiol oxidase
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T03788
R;Kiefer-Meyer, M.; Gomord, V.; O'Connell, A.; Halpin, C.; Faye, L.
Gene 178, 205-207, 1996
A;Title: Cloning and sequence analysis of laccase-encoding cDNA clones from tobacco.
A;Reference number: JC5229; MUID:97080572
A;Accession: T03788
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-409 <KIE>
A;Cross-references: EMBL:U45243; NID:g1685090; PIDN:AAC49538.1; PID:g1685091
C;Superfamily: laccase
C;Keywords: copper; oxidoreductase

Query Match 5.7%; Score 82; DB 2; Length 409;
Best Local Similarity 22.6%; Pred. No. 11;
Matches 47; Conservative 24; Mismatches 73; Indels 64; Gaps 8;

```
Qy     63 MAAAWP--SGPSAPEAVTAR-------LVGVLWFVSVTTGPWGAVATSAGGEESLKCEDL 113
          | || |    | | : :||         :|       :| |   | : :    ::||   |
Db    132 MVAASPFMDAPIAVDNITATATLHYSGALGTSPTTLTSTPPQNATSVANNFLDALK--SL 189

Qy    114 KVGQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPI 173
             :|  | |:  |      :   :|| :   | |  ||| ::|:    : | | |  |
Db    190 NSKKYPAKVPQTVD------HSLFFTAGLGINPCP--TCKQANGSRVVASVNNVTFVMP- 241

Qy    174 SCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQ 233
                  ||:  : | |  |   |   :||
Db    241 ----------TVALLQAHFFGINGV--FTTDFPA------------------------- 263

Qy    234 IVGPSDGSSYIIDYYGTRLTRLSITNET 261
                  :: :| ||  | |: || |
```

Db 263 ------NPPFVFNYTGTPPTNLATTNGT 284

RESULT 12
A54080
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 23-Jul-1999
C;Accession: A54080; I50592
R;Fang, K.S.; Barker, K.; Sudol, M.; Hanafusa, H.
J. Biol. Chem. 269, 14056-14063, 1994
A;Title: A transmembrane protein-tyrosine phosphatase contains spectrin-like repeats
A;Reference number: A54080; MUID:94245724
A;Accession: A54080
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1237 <FAN>
A;Cross-references: EMBL:Z21960; NID:g510510; PIDN:CAA79972.1; PID:g510511; GB:L13285
C;Superfamily: leukocyte common antigen; leukocyte common antigen cytosolic domain ho
C;Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phospha
F;528-1170/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F;786/Active site: Cys (phosphocysteine intermediate) #status predicted
F;792/Binding site: substrate phosphate (Arg) #status predicted

Query Match 5.6%; Score 80.5; DB 2; Length 1237;
Best Local Similarity 21.6%; Pred. No. 50;
Matches 48; Conservative 30; Mismatches 75; Indels 69; Gaps 11;

```
Qy     71 PSAPEAVTA----RLVGVLWFVSVTTGPWGAVATSAGG--------------EESLKCE 111
          | ||| :|     | | | |       | | :    |                | |  |
Db    323 PKAPENLTVHPTDRNVTVTWM-----KPTGTLEKHIDGYTVECNNTSQNVNRNETSFTCG 377

Qy    112 DL------------KVGQYICKD---PKINDATQ----EPVNCTNYTAHVSCFPAPNIT 151
          ||              |  :|  |:    |:| : |    :| | |::   :      :
Db    378 DLEPYSTGSVSVRAFKKSKYKNKNFEGEKVNGSFQTKPAKPENVTDFKLTLLADNTVKVA 437

Qy    152 CKDSS--GNETHF-------------TGNEVGFFKPISCRNVN---GYSYKVAVALSLF 192
          |:     |||| |               ||  |    : |:::    |::|::|   ::
Db    438 CRSQKVYGNETKFKLSWNSSSNSGENQRKNECNF----TVRDLSYLTKYTFKISVFNGVY 493

Qy    193 LGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQI 234
           |    :  |  | :  |: |    |  : : |  :|:  :|
Db    494 TGDSVCEEIYTRYNSRALIIFLV--FLIVVTSIALLLVLYKI 533
```

RESULT 13
LIPG
triacylglycerol lipase (EC 3.1.1.3) - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 30-Nov-1980 #sequence_revision 22-May-1981 #text_change 24-Feb-1995
C;Accession: A90638; A91256; A90675; A91124; A00732
R;De Caro, J.; Boudouard, M.; Bonicel, J.; Guidoni, A.; Desnuelle, P.; Rovery, M.
Biochim. Biophys. Acta 671, 129-138, 1981
A;Title: Porcine pancreatic lipase. Completion of the primary structure.
A;Reference number: A90638; MUID:82113655
A;Accession: A90638
A;Molecule type: protein
A;Residues: 308-449 <DEC>
A;Experimental source: pancreas
R;Bianchetta, J.D.; Bidaud, J.; Guidoni, A.A.; Bonicel, J.J.; Rovery, M.
Eur. J. Biochem. 97, 395-405, 1979
A;Title: Porcine pancreatic lipase. Sequence of the first 234 amino acids of the pept
A;Reference number: A91256; MUID:79236335
A;Contents: carbohydrate-binding site
A;Accession: A91256
A;Molecule type: protein
A;Residues: 1-234 <BIA>
A;Experimental source: pancreas
R;Guidoni, A.; Bonicel, J.; Bianchetta, J.; Rovery, M.
Biochimie 61, 841-845, 1979
A;Title: Porcine pancreatic lipase. Sequence between the 235th and 307th amino acids.
A;Reference number: A90675; MUID:80088446

A;Accession: A19947
A;Molecule type: DNA
A;Residues: 346-764 <CAM>
A;Cross-references: GB:J00125
A;Accession: B19947
A;Molecule type: mRNA
A;Residues: 339-509 <CA1>
A;Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536
R;Wu, L.; Morley, B.J.; Campbell, R.D.
Cell 48, 331-342, 1987
A;Title: Cell-specific expression of the human complement protein factor B gene: evidenc
A;Reference number: A25971; MUID:87102880
A;Accession: B25971
A;Molecule type: DNA
A;Residues: 1-99 <WUL>
A;Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534
R;Niemann, M.A.; Bhown, A.S.; Miller, E.J.
Biochem. J. 274, 473-480, 1991
A;Title: The principal site of glycation of human complement Factor B.
A;Reference number: S14339; MUID:91174758
A;Accession: S14339
A;Molecule type: protein
A;Residues: 270-329 <NIE>
A;Note: binding site for carbohydrate to lysine under artificial conditions
R;Morley, B.J.; Campbell, R.D.
EMBO J. 3, 153-157, 1984
A;Title: Internal homologies of the Ba fragment from human complement component factor B
A;Reference number: A44628; MUID:84158524
A;Accession: A44628
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 16-225,'F',227-259 <MOR>
R;Schwaeble, W.; Luttig, B.; Sokolowski, T.; Estaller, C.; Weiss, E.H.; Meyer zum Busche
Immunobiology 188, 221-232, 1993
A;Title: Human complement factor B: functional properties of a recombinant zymogen of th
A;Reference number: I54409; MUID:94041399
A;Accession: I54409
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-764 <RES>
A;Cross-references: GB:S67310; NID:g452937
R;Horiuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.
Mol. Immunol. 30, 1587-1592, 1993
A;Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conv
A;Reference number: I57824; MUID:94067177
A;Accession: I57824
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-31,'Q',33-764 <RE2>
A;Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922
C;Comment: 292-Cys has a free sulfhydryl.
C;Genetics:
A;Gene: GDB:BF
A;Cross-references: GDB:119726; OMIM:138470
A;Map position: 6p21.3-6p21.3
A;Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 69
A;Note: the list of introns may be incomplete
A;Note: gene is located in the major histocompatibility complex, class III region
C;Complex: complement factor B initially forms an inactive complex with complement factc
ment factor C3b forming active C3/C5 convertase; Ba is released
C;Function:
A;Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha c
al
A;Pathway: complement alternate pathway
C;Superfamily: complement C2; complement factor H repeat homology; trypsin homology
C;Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydrol
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-764/Product: complement factor B #status experimental <MAT>
F;26-259/Product: complement factor Ba fragment #status experimental <BAF>
F;37-98/Domain: complement factor H repeat homology <FH1>
F;103-158/Domain: complement factor H repeat homology <FH2>
F;165-218/Domain: complement factor H repeat homology <FH3>
F;260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>
F;482-752/Domain: trypsin homology #status atypical <TRY>
F;37-76,62-98,103-145,131-158,165-205,191-218,478-596,511-527,599-615,656-682,695-725
F;122,142,285,378/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental
F;526,576,699/Active site: His, Asp, Ser #status experimental

Query Match 5.8%; Score 83; DB 1; Length 764;
Best Local Similarity 24.1%; Pred. No. 18;
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

```
Qy     24 GTGLYPMRGPFKNLALLPFSLPLLGGG------------GSGSGEKVSV----------- 61
          |: | |       | |:|| | || ||            || | | | :
Db      2 GSNLSP------QLCLMPFILGLLSGGVTTTPWSLARPQGSCSLEGVEIKGGSFRLLQEG 55

Qy     61 SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPWGAVATS------AGGEESLKC-- 111
            :    |||  |   :  | |          :|| |  : |             :: |
Db     56 QALEYVCPSGFYPYPVQTRTCR----------STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy    111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
           | : |:|  : |  | : :   :|   ||    |      | ||: :     :|    |
Db    106 PHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy    167 VGFFK----PISCRNVNGYSYKV 185
           |:      ||  | | |  |::
Db    162 AGYCSNPGIPIGTRKV-GSQYRL 183
```

RESULT 9
T10729
transferrin-like protein Ttf-1, salt-induced - green alga (Dunaliella salina)
C;Species: Dunaliella salina
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: T10729
R;Fisher, M.; Gokhman, I.; Pick, U.; Zamir, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z17101
A;Accession: T10729
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1274 <FIS>
A;Cross-references: EMBL:U77059; NID:g1684791; PID:g1684792
C;Genetics:
A;Gene: ttf1
C;Superfamily: transferrin repeat homology

Query Match 5.8%; Score 83; DB 2; Length 1274;
Best Local Similarity 22.9%; Pred. No. 31;
Matches 30; Conservative 17; Mismatches 50; Indels 34; Gaps 6;

```
Qy     57 KVSVSKMAAAWPSGPSAPEAV-TARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKV 115
          :|    :   |    || :     |:|         | :| |  : ||
Db    579 QVDAETIEKFWEDNVCAPGSTENGPLIG--------GGKYGEVGENGGG---------- 620

Qy    116 GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISC 175
             :||  | :  :::|     |   | |         |  ||:  ||| :: | |  :
Db    620 ---LCKRCKTDCTSEDPY--AGYDGAVHCI-------DDDDGNQ--FTGGDIAFVKHSTL 665

Qy    176 RNVNGYSYKVA 186
          |: || :   |
Db    666 RDYNGPNLNTA 676
```

RESULT 10
JC4776
limbic-system-associated membrane protein precursor - human
C;Species: Homo sapiens (man)
C;Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 13-Nov-1998
C;Accession: JC4776
R;Pimenta, A.F.; Fischer, I.; Levitt, P.
Gene 170, 189-195, 1996
A;Title: cDNA cloning and structural analysis of the human limbic-system-associated m

```
Qy    166 EVGFFKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLKFCTVGFCGIG 222
Db    266 VVLVGLPRDAKCKSDVFTQVVKSVSIVGSYVGNRADTREALDFFARGLV-HAPIKIVGLS 324

Qy    223 SLIDFI--LISMQIVGPSDGSSYIID 246
Db    325 ELADVYDKMVKGEIVG-----RYVVD 345


RESULT   6
C70574
probable aroP2 protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Aug-1999
C;Accession: C70574
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
[redacted]re 393, 537-544, 1998
[redacted]thors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: C70574
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-487 <COL>
A;Cross-references: GB:Z95324; GB:AL123456; NID:g3261760; PIDN:CAB08578.1; PID:e315461;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: aroP2
C;Superfamily: arginine permease


  Query Match             5.8%;  Score 83.5;  DB 2;  Length 487;
  Best Local Similarity   26.5%;  Pred. No. 9.6;
  Matches   27;  Conservative   16;  Mismatches   40;  Indels   19;  Gaps    5;

Qy    159 ETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGF 218
Db      8 DERLTREDTGYHKGLHSRQLQMIALGGAIGTGLFLG--AGGRLASAGPGL----FLVYGI 61

Qy    219 CGIGSLIDFILISMQIVG------PSDGS--SYIIDYYGTRL 252
Db     62 CGI-----FVFLILRALGELVLHRPSSGSFVSYAREFYGEKV 98


[redacted]LT   7
[redacted]21
[redacted]ohol dehydrogenase (EC 1.1.1.1) 1 - yeast (Kluyveromyces marxianus var. marxianus)
C;Species: Kluyveromyces marxianus var. marxianus, Candida kefyr
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S32521
R;Ladriere, J.M.; Delcour, J.; Vandenhaute, J.
Biochim. Biophys. Acta 1173, 99-101, 1993
A;Title: Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase from Kluyverc
A;Reference number: S32521; MUID:93250057
A;Accession: S32521
A;Molecule type: DNA
A;Residues: 1-348 <LAD>
A;Cross-references: EMBL:X60224; NID:g297907; PIDN:CAA42785.1; PID:g297908
C;Genetics:
A;Gene: ADH1
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc
F;29-336/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F;98,101,104,112/Binding site: zinc, noncatalytic (Cys) #status predicted


  Query Match             5.8%;  Score 83;  DB 1;  Length 348;
  Best Local Similarity   20.8%;  Pred. No. 7.3;
  Matches   65;  Conservative   38;  Mismatches  107;  Indels  102;  Gaps   14;

Qy     19 NTRRDG---TGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSG----- 71
Db     37 NVKYSGVCHTDLHAWQGDWP----LDTKLPLV-GGHEGAGIVVAMGENVTGWEIGDYAGI 91

Qy     71 ------------------PSAPEA--------------VTARLV---------------- 83
Db     92 KWLNGSCMSCEECELSNEPNCPKADLSGYTHDGSFQQYATADAVQAARIPKNVDLAEVAP 151

Qy     83 ----GV-----LWFVSVTTGPWGAVATSAGGEESLKCEDLKV------------------ 116
Db    152 ILCAGVTVYKALKSAHIKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDAGDEKAKLF 211

Qy    116 ----GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE--VGF 169
Db    212 KELGGEYFIDFTKTKDMVAEVIEATNGVAHAVINVSVSEAAISTSVLYTRSNGTVVLVGL 271

Qy    170 FKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLK--FCTVGFCGIGSL 224
Db    272 PRDAQCK--SDVFNQVVKSISIVGSYVGNRADTREALDFFSRGLVKAPIKILGLSELASV 329

Qy    225 IDFILISMQIVG 236
Db    330 YD-KMVKGQIVG 340


RESULT   8
BBHU
complement factor B precursor - human
N;Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; he
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragmen
C;Species: Homo sapiens (man)
C;Date: 19-Feb-1984 #sequence_revision 05-Aug-1994 #text_change 18-Jun-1999
C;Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628;
R;Mejia, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.
submitted to the EMBL Data Library, March 1993
A;Reference number: S34075
A;Accession: S34075
A;Molecule type: mRNA
A;Residues: 1-764 <MEJ>
A;Cross-references: EMBL:X72875; NID:g297568; PIDN:CAA51389.1; PID:g297569
R;Woods, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.
Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982
A;Title: Isolation of cDNA clones for the human complement protein factor B, a class
A;Reference number: A44622; MUID:83039428
A;Accession: A44622
A;Molecule type: mRNA
A;Residues: 467-546;550-595;752-764 <WOO>
A;Cross-references: GB:J00185; GB:J00186
A;Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translat
R;Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.
J. Biol. Chem. 259, 3407-3412, 1984
A;Title: Complete primary structure for the zymogen of human complement factor B.
A;Reference number: A20751; MUID:84161997
A;Accession: A00934
A;Molecule type: protein; mRNA
A;Residues: 26-764 <MOL>
A;Cross-references: GB:K01566
A;Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 3
A;Note: 736-Ser was also found
A;Note: glycosylation sites were determined
R;Christie, D.L.; Gagnon, J.
Biochem. J. 209, 61-70, 1983
A;Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of
A;Reference number: A19188; MUID:83204002
A;Contents: the final paper in a series documenting the sequence, glycosylation site,
A;Accession: A19188
A;Molecule type: protein
A;Residues: 260-296,'T',298-764 <CHR>
R;Campbell, R.D.; Porter, R.R.
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983
A;Title: Molecular cloning and characterization of the gene coding for human compleme
A;Reference number: A19947; MUID:83273641
```

```
  Query Match             6.4%;  Score 92;  DB 2;  Length 456;
  Best Local Similarity   26.2%;  Pred. No. 1.7;
  Matches   42;  Conservative   11;  Mismatches   45;  Indels   62;  Gaps    7;

Qy      47 LGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEE 106
           ||||| | |:                        |:|                  |||
Db      13 LGGGGGGGGD------------------------LLG------------------GGEA 29

Qy     107 S----LKCEDLKVGQYIC-KDPKINDATQEPVNCTNYTAHV------SCFPAPNITCKDS 155
                 || |:||:  | :      |  :   |  : ||:      |||    :||  |
Db      30 DGLMRALCEGLRVGEEDCARFVLYGVAYWQGGRCPEWVAHITRCADLSCFATYLLTCHRS 89

Qy     156 SGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGW 195
           | |  |||  |   :  | |       | |  ||||  :
Db      90 GGCE--FTGGRVARDRLPSLRE------SVEVLQSLFLAF 121


RESULT   3
A59
lycoprotein precursor - murine hepatitis virus (strain A59)
N;Alternate names: peplomer glycoprotein; spike glycoprotein
C;Species: murine hepatitis virus, MHV
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 12-Apr-1996
C;Accession: A27402
R;Luytjes, W.; Sturman, L.S.; Bredenbeek, P.J.; Charite, J.; van der Zeijst, B.A.M.; Hor
Virology 161, 479-487, 1987
A;Title: Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identificat
A;Reference number: A27402; MUID:88072088
A;Accession: A27402
A;Molecule type: genomic RNA
A;Residues: 1-1324 <LUY>
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1324/Product: E2 glycoprotein #status predicted <E2G>
F;17-717/Product: 90B glycoprotein #status predicted <EGB>
F;718-1324/Product: 90A glycoprotein #status predicted <EGA>
F;1266-1286/Domain: transmembrane #status predicted <TMN>
F;31,60,192,247,357,435,442,530,625,657,665,688,737,754,844,893,1126,1180,1190,1209,1225

  Query Match             6.3%;  Score 91;  DB 1;  Length 1324;
  Best Local Similarity   23.6%;  Pred. No. 6.7;
  Matches   59;  Conservative   27;  Mismatches   94;  Indels   70;  Gaps   13;

Qy      25 TGLYPMRG-PFKNLALLP--------FSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSA-- 74
           || ||: |  |:|||           |  | |      |   |  |  :  : ||| :|
Db      66 TGYYPVDGSKFRNLALTGTNSVSLSWFQPPYLNQFNDGIFAK--VQNLKTSTPSGATAYF 123

Qy      74 PEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQY-ICKDPKINDATQEP 132
           |  |   | |   : :|   |:  |  ::           | || ||: |
Db     124 PTIVIGSLFGYTSY-TVVIEPYNGVIMAS-----------VCQYTICQLP--------- 162

Qy     133 VNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE-VGFF----KPISCRNVNGYSYKVAV 187
                :  | |  |               ||: :||:    ||  |      ::  |
Db     162 --------YTDCKPNTN-------------GNKLIGFWHTDVKPPICVLKRNFTLNVNA 199

Qy     188 ALSLFLGWLGADRFYLGY---PALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
             |  :    ||  |   |:     |       || ::    :   |  |: ||::
Db     200 DAFYFHFYQHGGTFYAYYADKPSATTFLFSVY----IGDILTQYYVLPFICNPTAGSTFA 255

Qy     245 IDYYGTRLTR 254
           |: | | :
Db     256 PRYWVTPLVK 265


RESULT   4
T08604
hypothetical protein GRR1 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08604
R;Chen, W.; Atherly, A.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z15438
A;Accession: T08604
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-690 <CHE>
A;Cross-references: EMBL:AF019910; NID:g2407789; PID:g2407790
A;Experimental source: variety L85-3044; root
C;Genetics:
A;Gene: grr1

  Query Match             6.2%;  Score 89.5;  DB 2;  Length 690;
  Best Local Similarity   24.6%;  Pred. No. 4.3;
  Matches   43;  Conservative   22;  Mismatches   59;  Indels   51;  Gaps    7;

Qy      85 LWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPV--NCTNYTAHV 142
           || |: | |  | :  ::|  :  | :       :|| | |:| |   |  || |  | :
Db     233 LWDVA-TVGDVGLIEIASGCHQLEKLD-------LCKCPNISDKTLIAVAKNCPN-LAEL 283

Qy     143 SCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFY 202
           |    |||      |||      :    : || :| :|   :
Db     284 SIESCPNI------GNEGLQAIGKCPNLRSISIKNCSGVGDQ----------------- 320

Qy     203 LGYPALGLLKFCTVGFCGIGSLIDFIL--ISMQIVGPSDGSSYIIDYYGTRLTRL 255
                       |  |: |   | |  : :: :  || |  :| :||  :| |
Db     320 -------------GVAGLLSSASFALTKVKLESLTVSDLSLAVIGHYGVAVTDL 360


RESULT   5
S20911
alcohol dehydrogenase (EC 1.1.1.1) II - yeast (Kluyveromyces marxianus var. lactis)
C;Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 11-Jun-1999
C;Accession: S20911; S19804
R;Shain, D.H.; Salvadore, C.; Denis, C.L.
Mol. Gen. Genet. 232, 479-488, 1992
A;Title: Evolution of the alcohol dehydrogenase (ADH) genes in yeast: characterizatio
A;Reference number: S20911; MUID:92269769
A;Accession: S20911
A;Molecule type: DNA
A;Residues: 1-348 <SHA>
A;Cross-references: EMBL:X64397; NID:g2832; PIDN:CAA45739.1; PID:g2833
C;Genetics:
A;Gene: ADH2
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc
F;29-336/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;173-202/Region: beta-alpha-beta NAD nucleotide-binding fold
F;44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F;98,101,104,112/Binding site: zinc, noncatalytic (Cys) #status predicted

  Query Match             5.9%;  Score 85.5;  DB 2;  Length 348;
  Best Local Similarity   19.9%;  Pred. No. 4.4;
  Matches   65;  Conservative   35;  Mismatches  111;  Indels  115;  Gaps   13;

Qy      19 NTRRDG---TGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSVSKMAAAW-------- 68
           | :  |   | |:  :| :     ||  |||: ||  |:|  |:: :    |
Db      37 NVKYSGVCHTDLHAWKGDWP----LPTKLPLV-GGHEGAGVVVAMGENVKGWIIGDFAGI 91

Qy      68 ----------------------------------------------PSGPSAPEAVT 78
                                                        | |    |
Db      92 KWLNGSCMSCEYCELSNESNCPDADLSGYTHDGSFQQYATADAVQAARIPKGTDLAEVAP 151

Qy      79 ARLVGV-----LWFVSVTTGPWGAVATSAGGEESLKCEDLKV----------------- 116
                ||      |    :  | | |:: : ||  ||  :  |
Db     152 ILCAGVTVYKALKSADLKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDTGAEKAKLF 211

Qy     116 ----GQYICKDPKINDATQEPVNCTNYTAH------VSCFPAPNITCKDSSGNETHFTGN 165
               |:|        |  :| |: ||  ||      || |        : | |     |
Db     212 KELGGEYFVDYAVSKDLIKEIVDATNGGAHGVINVSVSEFAI------EQSTNYVRSNGT 265
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 07:23:59 ; Search time 35.93 Seconds
(without alignments)
353.146 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 1439
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	167.5	11.6	573	2	S44605	C02F5.3 protein -
2	92	6.4	456	2	S55661	hypothetical prote
3	91	6.3	1324	1	VGIH59	E2 glycoprotein pr
4	89.5	6.2	690	2	T08604	hypothetical prote
5	85.5	5.9	348	2	S20911	alcohol dehydrogen
6	83.5	5.8	487	2	C70574	probable aroP2 pro
7	83	5.8	348	1	S32521	alcohol dehydrogen
8	83	5.8	764	1	BBHU	complement factor
9	83	5.8	1274	2	T10729	transferrin-like p
10	82.5	5.7	338	2	JC4776	limbic-system-asso
11	82	5.7	409	2	T03788	laccase (EC 1.10.3
12	80.5	5.6	1237	2	A54080	protein-tyrosine-p
13	80	5.6	449	1	LIPG	triacylglycerol li
14	80	5.6	328	1	A41927	insulin-like growt
15	80	5.6	171	2	T01466	nifU protein homol
16	79.5	5.5	4367	1	B54802	dynein heavy chain
17	79.5	5.5	3432	1	GNWVJE	genome polyprotein
18	79.5	5.5	3432	1	GNWVJS	genome polyprotein
19	79.5	5.5	489	2	B70514	probable integral
20	79.5	5.5	910	2	S68983	auxilin - bovine
21	79	5.5	525	2	S28052	sucrose transport
22	78.5	5.5	482	2	F71969	hypothetical prote
23	78.5	5.5	398	2	A35281	integumentary muci
24	78.5	5.5	7962	2	I38346	elastic titin - hu
25	78	5.4	557	2	H70725	hypothetical prote
26	78	5.4	523	2	S51114	sucrose-proton sym
27	77.5	5.4	2179	1	GNNYH4	genome polyprotein
28	77.5	5.4	374	2	A69200	chorismate synthas
29	77	5.4	1239	1	A32579	neuroglian - fruit
30	76.5	5.3	770	2	S37963	mitochondrial inte
31	76.5	5.3	559	2	S52956	period protein (cl
32	76.5	5.3	559	2	S52955	period protein (cl
33	76	5.3	482	1	A34671	triacylglycerol li
34	76	5.3	467	2	A43357	pancreatic lipase-
35	76	5.3	775	2	T00962	hypothetical prote
36	76	5.3	1556	2	D36793	hypothetical prote
37	76	5.3	237	2	A42013	alpha-1-B-glycopro
38	76	5.3	203	2	JC4871	phospholipase C (E
39	75.5	5.2	592	1	IKEBCA	colicin A - Citrob
40	75.5	5.2	350	2	S09475	alcohol dehydrogen
41	75.5	5.2	560	2	S52943	period protein (cl
42	75.5	5.2	1127	2	A25018	circadian rhythm p
43	75.5	5.2	1122	2	B26427	period clock prote
44	75.5	5.2	1218	2	A26427	period clock prote
45	75.5	5.2	1176	2	C26427	period clock prote

ALIGNMENTS

RESULT 1
S44605
C02F5.3 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 16-Jul-1999
C;Accession: S44605
R;Anderson, K.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the C. elegans cosmid C02F5.
A;Reference number: S44603
A;Accession: S44605
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-573 <AND>
A;Cross-references: EMBL:L14745; NID:g289607; PID:g289610
C;Genetics:
A;Introns: 224/2; 304/1; 363/3; 390/3; 503/2
C;Superfamily: translation elongation factor Tu homology
C;Keywords: GTP binding
F;63-183/Domain: translation elongation factor Tu homology <ETU>
F;69-76/Region: nucleotide-binding motif A (P-loop)
F;246-249/Region: GTP-binding NKXD motif

Query Match 11.6%; Score 167.5; DB 2; Length 573;
Best Local Similarity 27.9%; Pred. No. 7e-07;
Matches 50; Conservative 23; Mismatches 57; Indels 49; Gaps 5;

```
Qy      90 VTTGPWGAVATSAGGEESLKCEDLKVGQYICKDP----------------------K 124
           |:| | | |          ::|  |:    :|:||                      |
Db     415 VSTNPLGPV---------VECRFLENSFILCEDPVPLYGPGQTGQQPANESFRNEGKCLK 465

Qy     125 INDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYK 184
           :     | |  ||   | |   | | |    |  |        | |   |   ||: :
Db     466 MGGYRAEDVEFTN----VKCRVLPCIEC---HGPRT--------FTKSTPCIIYNGHYFL 510

Qy     185 VAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSY 243
             :  |:||| :  ||| ||| |: : |  |:|  ||  ::|  |: : ::||:| ||:
Db     511 TTLLYSIFLGVVAVDRFCLGYSAMAVGKLMTLGGFGIWWIVDIFLLVLGVLGPADDSSW 569
```

RESULT 2
S55661
hypothetical protein 66 - equine herpesvirus 2
C;Species: equine herpesvirus 2
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C;Accession: S55661
R;Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A;Title: The DNA sequence of equine herpesvirus 2.
A;Reference number: S55594; MUID:95302501
A;Accession: S55661
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-456 <TEL>
A;Cross-references: GB:U20824; NID:g695172; PID:g695239
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

```
PT   anaemia, AIDS and various cancers
PS   Disclosure; Page 29-30; 33pp; English.
CC   A human T-cell lambda-gt10 random primed cDNA library was
CC   screened with a fragment corresponding to the extracellular
CC   domain of mouse flt3 ligand (flt3-L) (nt 103-516 of Q79076)
CC   to isolate human flt3-L cDNA.  Flt-3 stimulates progenitor and
CC   stem cells, and can be used e.g. in gene therapy protocols.
SQ   Sequence   988 BP;    184 A;    360 C;    269 G;    175 T;


  Query Match             4.0%;  Score 32.8;  DB 1;  Length 988;
  Best Local Similarity   56.5%;  Pred. No. 2.7;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    127 TCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAG 186
          |||| |    ||| | ||  ||  | |   | | | || || |   ||| || ||||| |
Db    108 TCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGATAGGTTGTTGGGCTCCAGG 49

Qy    187 ATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACG 234
          ||||| | ||   | | | | | |||   || |||||   |  | ||
Db     48 CTGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAATTCCGGCCG 1


RESULT  14
V81506/c
ID   V81506 standard; cDNA to mRNA; 988 BP.
AC   V81506;
DT   25-MAR-1999  (first entry)
DE   Human flt3-ligand encoding cDNA.
KW   Antigen-specific peripheral immune tolerance; flt3-ligand; flt3-L;
KW   immunogenic; autoimmune disease; organ transplantation; food allergy;
KW   tissue transplantation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             30. .737
FT                   /*tag=  a
PN   WO9857655-A1.
PD   23-DEC-1998.
PF   12-JUN-1998; U12085.
PR   17-JUN-1997; US-877421.
PA   (IMMV ) IMMUNEX CORP.
PI   Abbott NM, Mowat AM, Viney JL;
DR   WPI; 99-070422/06.
DR   P-PSDB; W67769.
PT   Methods for initiating or enhancing antigen specific immune
PT   tolerance  - by using murine or human flt3 ligand
PS   Disclosure; Page 13-14; 25pp; English.
CC   A method has been developed of initiating or enhancing: (i) an antigen-
CC   specific immune tolerance; or (ii) immunotolerance of a therapeutic
CC   immunogenic molecule by addition of a polypeptide, before, after or with
CC   the mucosal administration of an immunotolerising amount of the antigen
CC   or therapeutic molecule, respectively. The polypeptide is capable of
CC   binding the flt3 receptor and is: a) amino acids 28-x of murine flt3
CC   ligand (flt3-L), where x is an amino acid between 163-231; b) amino
CC   acids 28-y of human flt3-L, where y is an amino acid between 160-235;
CC   and c) a polypeptide that has at least 90% identity to the polypeptides
CC   of either (a) or (b). The method ameliorates the effects of autoimmune
CC   diseases, food allergies or organ or tissue rejection following
CC   transplantation. Administration of flt3-L allows lower doses of antigens
CC   to be used in vivo for mucosally administered antigens. The present
CC   sequence encodes human flt3-L.
SQ   Sequence   988 BP;    184 A;    360 C;    269 G;    175 T;


  Query Match             4.0%;  Score 32.8;  DB 1;  Length 988;
  Best Local Similarity   56.5%;  Pred. No. 2.7;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    127 TCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCTCCAAG 186
          |||| |    ||| | ||  ||  | |   | | | || || |   ||| || ||||| |
Db    108 TCCCACTGAGTCCCGAGCTCAGCAGCAGCAGGAGGAGATAGGTTGTTGGGCTCCAGG 49

Qy    187 ATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACG 234
          ||||| | ||   | | | | | |||   || |||||   |  | ||
Db     48 CTGGCGCCAGCACTGTCATTTCGGCCGGGGGCCCCGGAATTCCGGCCG 1


RESULT  15
X00726
ID   X00726 standard; cDNA to mRNA; 3250 BP.
AC   X00726;
DT   30-MAR-1999  (first entry)
DE   Human aggrecan degrading metalloprotease 2 gene.
KW   Human; aggrecan degrading metalloprotease; cartilage; proteoglycan;
KW   interglobular domain; matrix metalloprotease; bovine; interleukin-1beta;
KW   primer; PCR; amplification; inhibitor; cleavage; inhibitor; ADMP; ss;
KW   arthritis; joint injury; pseudogout.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             121. .2913
FT                   /*tag=  a
FT                   /product= ADMP-2
FT                   /note= "aggrecan degrading metalloprotease 2"
PN   WO9905291-A2.
PD   04-FEB-1999.
PF   24-JUL-1998; U15438.
PR   16-OCT-1997; US-062169.
PR   25-JUL-1997; US-053850.
PR   15-AUG-1997; US-055836.
PA   (DUPO ) DU PONT PHARM CO.
PI   Arner EC, Burn TC, Copeland RA, Decicco CP, Liu R, Magolda R,
PI   Pratta M, Solomon KA, Tortorella MD, Trzaskos JM, Yang F;
DR   WPI; 99-142943/12.
DR   P-PSDB; W75426.
PT   New isolated aggrecan degrading metallo proteases - used to develop
PT   products for treating, e.g. osteoarthritis, joint injury, reactive
PT   arthritis, psoriatic arthritis or juvenile rheumatoid arthritis
PS   Claim 17; Page 65-68; 73pp; English.
CC   This sequence represents the coding region for the human aggrecan
CC   degrading metalloprotease 2 (ADMP-2).  ADMP-1 (X00725) and ADMP-2 are
CC   novel proteases that cleave the aggrecan (a major cartilage proteoglycan)
CC   between residues Glu373-Ala374 of the interglobular domain (compared with
CC   cleavage between Asn341-Phe342 by the matrix metalloproteases MMP-1, -2,
CC   -3, -7, -8, -9 and -13).  ADMP-1 and ADMP-2 were isolated and purified
CC   from the conditioned media of bovine nasal cartilage stimulated by
CC   interleukin-1beta.  The purified proteins were partially sequenced and
CC   primers were synthesised based on the resultant amino acid sequences
CC   (X00727-X00732 for ADMP-1 and X00733-X00736 for ADMP-2).  These were
CC   used to isolate the corresponding genes from cDNA.  The ADMP polypeptides
CC   can be used for identifying inhibitors of ADMP activity which would
CC   prevent cleavage of the aggrecan core protein, thereby decreasing the
CC   loss of aggrecan from cartilage. Such inhibitors can be used for treating
CC   diseases such as osteoarthritis, joint injury, reactive arthritis, acute
CC   pyrophosphate arthritis (pseudogout), psoriatic arthritis and juvenile
CC   rheumatoid arthritis.
SQ   Sequence   3250 BP;    761 A;    900 C;    919 G;    670 T;


  Query Match             4.0%;  Score 32.8;  DB 1;  Length 3250;
  Best Local Similarity   56.5%;  Pred. No. 5.1;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    122 CCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTCT 181
          ||  | ||| |  || |   |||| ||| | ||| ||  |  ||   | |  | |   ||
Db    300 CCCGCACCCCCTGGCGCAGCGGCGCAGGAGCAAGGGGCTGGTGCAGAACATCGACCAACT 359

Qy    182 CCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCG 229
          | |    |||||| ||  ||  | ||    ||||| | | || |||||
Db    360 CTACTCCGGCGGCGGCAAGGTGGGCTACCTCGTCTACGCGGGCGGCCG 407


Search completed: February 11, 2000, 05:26:34
Job time: 186 sec
```

```
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    5401. .5460
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    7201. .7260
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    9001. .9060
FT                   /*tag= f
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 594-599; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   9179 BP;    2976 A;    1267 C;    1624 G;    2950 T;

  Query Match             4.1%;  Score 33.4;  DB 1;  Length 9179;
  Best Local Similarity   51.7%;  Pred. No. 5.8;
  Matches   76;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     655 TGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCA 714
           | ||  ||| ||    | |||  || | |  ||  ||  |       | || |||  ||
Db    4554 TTTGCGATTCGGTTGATGATTTGTTACGTCGTTCATTTCAGTTGTCATATTAGACACTCC 4613

Qy     715 GATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACT 774
           || || | || | |||||   | |||||   |  | |    ||| | | |  ||   |
Db    4614 TTTGAAAATTGTAATATTATCTTTAACTATAACAAAATATAATCAAAAATAAACATGTTT 4673

Qy     775 AATGAAACATTTAGAAAAACGCAATTA 801
           | | ||  |||   |||||   || ||
Db    4674 ATTAAACAATTATTAAAAATAAAAATA 4700


RESULT  12
V90286/c
ID   V90286 standard; cDNA; 591 BP.
AC   V90286;
DT   15-FEB-1999  (first entry)
DE   EST clone DJ238.
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
OS   Homo sapiens.
PN   WO9845436-A2.
PD   15-OCT-1998.
PF   10-APR-1998; U06955.
PR   10-APR-1997; US-838821.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070077/06.
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
PS   Claim 1; Page 491; 618pp; English.
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
SQ   Sequence   591 BP;    149 A;    134 C;    135 G;    173 T;

  Query Match             4.1%;  Score 33;  DB 1;  Length 591;
  Best Local Similarity   55.8%;  Pred. No. 1.8;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     692 CAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACCA 751
           ||| |||| ||| || ||   | | ||| |  | |||| ||    ||  |  | |||||
Db     348 CAAGGCAGTTTGCTGCACAGACTGTTGGCAACACTTACGGTAATTTTTCGTTAGCAACCA 289

Qy     752 GACTTACAAGACTGAGTATTACTAATGAAACATTTAGAAAAACGCAATTATAT 804
           || | ||   |    |||| || |||   | || ||| | |  | |  ||
Db     288 TGTTTCCCAGGAGGGAATTTACCAAAGAAGATTATAAAAAGAAGTTACTGGAT 236


RESULT  13
Q79079/c
ID   Q79079 standard; cDNA to mRNA; 988 BP.
AC   Q79079;
DT   05-AUG-1995  (first entry)
DE   Human flt-3 ligand cDNA.
KW   Flt-3 ligand; flt3-L; anemia; cancer; AIDS; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds              30. .737
FT                    /*tag=  a
PN   EP-627487-A.
PD   07-DEC-1994.
PF   19-MAY-1994; 303575.
PR   24-MAY-1993; US-068394.
PR   12-AUG-1993; US-106463.
PR   25-AUG-1993; US-111758.
PR   03-DEC-1993; US-162407.
PR   07-MAR-1994; US-209502.
PR   11-MAY-1994; US-243545.
PA   (IMMV ) IMMUNEX CORP.
PI   Beckmann MP,  Lyman SD;
DR   WPI; 95-008071/02.
DR   P-PSDB; R67541.
PT   Isolated ligands for flt 3 receptors - useful for treating
```

```
PT   Tomato polygalacturonase gene -
PT   used for modulating expression in plant cells or directing
PT   expression of hetrologous peptide(s).
PS   Disclosure; ; English.
CC   cDNA of gene may act as a probe to the genomic sequence, provides a means
CC   of modulating the production of PG and acts as a source of the transit
CC   peptide which may be joined to heterologous peptides directing them to
CC   cell wall.
SQ   Sequence   1624 BP;   586 A;   231 C;   294 G;   513 T;


  Query Match             4.1%;  Score 33.4;  DB 1;  Length 1624;
  Best Local Similarity   58.6%;  Pred. No. 2.3;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    687 TATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGG 746
          ||||||| |  ||||| ||||| | | ||| | |  |||| | |||  ||| ||| | |
Db    445 TATTTCAGTAAAGATTTTTGGATCCTTAGAAGCATCTAGTAAAATTTCAGACTACAAAGA 504

Qy    747 AACCAGACTTACAAGACTGAGTATTACTAATGAAACATT 785
          |  || |||   |       |  || |  | |||  ||
Db    505 TAGAAGGCTTTGGATTGCTTTTGATAGTGTTCAAAATTT 543


RESULT   9
N80487
ID   N80487 standard; DNA; 1637 BP.
AC   N80487;
DT   20-NOV-1990  (first entry)
DE   Plasmid clone pTom6.
KW   Fruit ripening; polygalacturonase; pectin esterase; ss.
FH   Key             Location/Qualifiers
FT   cds             51. .1424
FT                   /*tag=  a
FT                   /label=polygalacturonase
PN   EP-271988-A.
PD   22-JUN-1988.
PF   06-NOV-1987; 309853.
PR   11-NOV-1986; GB-026879.
PA   (ICIL) Imperial Chem Inds PLC.
PI   Bridges IG, Schuch WW, Grierson D.;
DR   WPI; 88-169271/25.
DR   P-PSDB; P80299.
PT   Recombinant DNA comprising promoter and terminator sequences -
PT   useful in plants for altering ripening properties esp. in tomatoes.
PS   Disclosure; p; English.
CC   This cDNA clone, encoding polygalacturonase (PG), is used to produce
CC   antisense mRNA (with an inverted sequence to that of PG mRNA) which is
CC   inserted into a vector used to transform plants which thereafter have
CC   altered ripening properties.  The inverted sequence and the PG mRNA
CC   form a double-stranded structure which inhibits expression of the PG
CC   mRNA.  See also N80488.
SQ   Sequence   1637 BP;    598 A;    232 C;    290 G;    517 T;


  Query Match             4.1%;  Score 33.4;  DB 1;  Length 1637;
  Best Local Similarity   58.6%;  Pred. No. 2.3;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    687 TATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGG 746
          ||||||| |  ||||| ||||| | | ||| | |  |||| | |||  ||| ||| | |
Db    449 TATTTCAGTAAAGATTTTTGGATCCTTAGAAGCATCTAGTAAAATTTCAGACTACAAAGA 508

Qy    747 AACCAGACTTACAAGACTGAGTATTACTAATGAAACATT 785
          |  || |||   |       |  || |  | |||  ||
Db    509 TAGAAGGCTTTGGATTGCTTTTGATAGTGTTCAAAATTT 547


RESULT  10
Q38415
ID   Q38415 standard; DNA; 1636 BP.
AC   Q38415;
DT   05-JUL-1993  (first entry)
DE   Polygalacturonase cDNA clone pTOM6.
KW   polygalacturonase; PG; pectin esterase; PE; expression regulation;
KW   fruit softening enzymes; flowering plants; fruiting plants;
KW   antisense RNA.
OS   Lycospersicon esculentum.
FH   Key             Location/Qualifiers
FT   cds             51. .1422
FT                   /*tag=  a
FT                   /label= polygalacturonase (PG)
PN   EP-532060-A.
PD   17-MAR-1993.
PF   06-NOV-1987; 117411.
PR   11-NOV-1986; GB-026879.
PA   (ICIL ) IMPERIAL CHEM IND PLC.
PI   Bridges IG,  Grierson D,  Schuch WW;
DR   WPI; 93-087084/11.
DR   P-PSDB; R32107.
PT   Recombinant DNA for flowering and fruiting plants e.g. tomatoes
PT   ripening control - comprises base sequence for transcription
PT   contg. inverted sequence of bases complementary to bases in anti
PT   sense ribonucleic acid encoding softening enzymes, or gene
PT   expression regulation
PS   Example 12; Fig 1; 20pp; English.
CC   This is the sequence of clone pTOM6 which contains the coding
CC   sequence of polygalacturonase (PG). It was used to isolate the
CC   PG promoter in the construction of a vector encoding antisense RNA
CC   to the PG cDNA and PG gene. This would be useful to regulate the
CC   expression of the fruit softening enzymes in flowering and fruiting
CC   plants. Such antisense RNA would delay fruit softening.
SQ   Sequence   1636 BP;    598 A;    232 C;    290 G;    516 T;


  Query Match             4.1%;  Score 33.4;  DB 1;  Length 1636;
  Best Local Similarity   58.6%;  Pred. No. 2.3;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    687 TATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGG 746
          ||||||| |  ||||| ||||| | | ||| | |  |||| | |||  ||| ||| | |
Db    449 TATTTCAGTAAAGATTTTTGGATCCTTAGAAGCATCTAGTAAAATTTCAGACTACAAAGA 508

Qy    747 AACCAGACTTACAAGACTGAGTATTACTAATGAAACATT 785
          |  || |||   |       |  || |  | |||  ||
Db    509 TAGAAGGCTTTGGATTGCTTTTGATAGTGTTCAAAATTT 547


RESULT  11
V74411
ID   V74411 standard; DNA; 9179 BP.
AC   V74411;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #100.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    1. .60
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
FT   misc_feature    1801. .1860
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
FT   misc_feature    3601. .3660
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
```

```
PT   basic-helix-loop-helix genes - especially for identifying
PT   medulla:blastoma and assessing its aggressiveness, specifically
PT   associated with expression of BHLH genes neuroD 1-3
PS   Example 11; Columns 69-72; 43pp; English.
CC   The present sequence encodes a protein which is a member of the basic
CC   helix-loop-helix (bHLH) protein family, and is designated neuroD2. The
CC   bHLH genes are a family of genes associated with vertebrate neuronal,
CC   endocrinal and gastrointestinal development. The observed pattern of
CC   neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC   The specification describes a method for the classification of human
CC   neuroectodermal tumours. The method comprises measuring, in a tumour
CC   sample, expression of at least one basic bHLH gene and identifying the
CC   tumour subclass by matching expression to predetermined expression
CC   profiles for known subclasses. For classifying the tumour as a
CC   medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC   The method is used to classify neuroectodermal tumours, and to identify
CC   medullablastoma and for prognosis of this as aggressive.
SQ   Sequence   1951 BP;    359 A;    668 C;    557 G;    353 T;

  Query Match             4.5%;  Score 36.4;  DB 1;  Length 1951;
  Best Local Similarity   49.0%;  Pred. No. 0.31;
  Matches   97;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180
          | | ||||   |  || |  |  ||   || |  |||   | | |||  | | | ||
Db    260 CTCCTCTCGGACGTGCCCAAGTTCGCCAGCTGGGGCGACGGCGACGACGACGAGCCGAGG 319

Qy    181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
          | | | ||  | |||   ||||     | |  |||||||| | | | | | |  |  ||
Db    320 AGCGACAAGGGCGACGCGCCGCCGCAGCCTTCTCCTGCTCCCGGGTCGGGGGCTCCAGGA 379

Qy    241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300
          | ||   | | |     |   || ||  |   |  ||||     |    |||   ||||
Db    380 CCCGCCCGGGCCGCCAAGCCAGTGTCTCTTCGTGGAGGAGAAGAGATCCCTGAACCCACG 439

Qy    301 TCCGCCGGGGGCGAGGAG 318
          |  || | || | |||||
Db    440 TTGGCTGAGGTCAAGGAG 457


RESULT   6
T58840_3
Continuation (4 of 6) of T58840 from base 300001 (Mycoplasma genitalium genome. )
WP Sequence split into 6 fragments  LOCUS T58840 Accession T58840
WP     Fragment Name      Begin     End
WP     T58840_0               1     110000
WP     T58840_1          100001     210000
WP     T58840_2          200001     310000
WP     T58840_3          300001     410000
WP     T58840_4          400001     510000
WP     T58840_5          500001     580073


  Query Match             4.4%;  Score 35.4;  DB 1;  Length 110000;
  Best Local Similarity   55.2%;  Pred. No. 5.4;
  Matches   69;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy    685 CTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTAT 744
          ||| ||| ||   | |||||    | ||  ||    |       ||  || || |  |
Db  87060 CTTTTTTAAAGTTAAATTGTCCTGCTTTGTGAATCCATGGTAAGCAGGATGGAATTATTA 87119

Qy    745 GGAACCAGACTTACAAGACTGAGTATTACTAATGAAACATTTAGAAAAACGCAATTATAT 804
           | |||   | | |||| || | ||| | | || ||||| | || |||    | ||| |
Db  87120 AGTACCTTCCATCCAAGGCTTATTATGAATCATTAAACAATAAGCAAATTCAATTTAAAG 87179

Qy    805 CCATA 809
          |||||
Db  87180 CCATA 87184


RESULT   7
Q27189/c
ID   Q27189 standard;. DNA; 4673 BP.
AC   Q27189;
DT   01-FEB-1993  (first entry)
DE   P. yoelii SSP2 antigen.
KW   Sporozoite; surface protein; NYS4; vaccine; malaria; CSP;
KW   circumsporozoite; ss.
OS   Plasmodium yoelii.
FH   Key             Location/Qualifiers
FT   cds             718. .3198
FT                   /*tag= a
FT                   /label= SSP2
PN   WO9211868-A.
PD   23-JUL-1992.
PF   03-JAN-1992; U00018.
PR   10-JAN-1991; US-638431.
PA   (USNA ) US SEC OF NAVY.
PI   Charoenvit Y, Hedstrom R, Hoffman SL, Khusmith S, Rogers W;
PI   Rogers WO;
DR   WPI; 92-284327/34.
DR   P-PSDB; R26042.
PT   Protective malaria sporozoite surface protein immunogen and gene
PT   - for vaccination preferably in combination with circumsporozoite
PT   protein
PS   Claim 1; Page 38  and pages 16-30; 43pp; English.
CC   SSP2 on the surface of Plasmodium sporozoites is detected by a
CC   monoclonal antibody (NYS4) which is specific for a 140 kD protein.
CC   SSP2 includes a short hydrophobic leader sequence, a region contg.
CC   perfect or degenerate tandem repeats of short peptide motifs, flanking
CC   regions contg. non-repetitive amino acid sequence, a membrane spanning
CC   domain and a cytoplasmic domain.
CC   The protein (or fragments of it) may be used to provide a vaccine
CC   against malaria in conjunction with other malarial antigens, e.g.
CC   circumsporozoite proteins of P. yeolii, P. falciparum, P. vivax, P.
CC   malariae and P. ovale.
SQ   Sequence   4673 BP;    1835 A;    807 C;    530 G;    1501 T;

  Query Match             4.3%;  Score 34.6;  DB 1;  Length 4673;
  Best Local Similarity   55.4%;  Pred. No. 1.8;
  Matches   67;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    560 TCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTG 619
          | |  |||| |   |||  ||| | ||| ||    |     |||    |||| |    ||
Db   1993 TTGGCTTGTTTGGGTTTGATGGTTCGTTTGGGTTTGGCTTGTTTGGGTTTGGCTTGTTTG 1934

Qy    620 CTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATT 679
          ||| || |||||   |||| ||  | | |||||| |     |||| |  |  |||  |
Db   1933 GTTTTGGCTTGTTTGGGTTTCGCCTTTTTGGGTTTCGCCTTTTTGGGTGGTTGTTTGGGT 1874

Qy    680 T 680
          |
Db   1873 T 1873


RESULT   8
N91112
ID   N91112 standard; DNA; 1624 BP.
AC   N91112;
DT   21-JUN-1990  (first entry)
DE   Polygalacturonase (PG) gene cDNA.
KW   Polygalacturonase gene; tomato; ds.
OS   Lycopersicon esculentum.
FH   Key             Location/Qualifiers
FT   cds             47. .1418
FT                   /*tag= a
PN   US4801540-A.
PD   31-JAN-1989.
PF   2-JAN-1987; 000201.
PR   2-JAN-1987; US-000201.
PA   (CALG-) Calgene Inc.
PI   Hiatt WR, Sheehy RE, Shewmaker CK, Kridl JC, Knauf V;
DR   WPI; 89-053640/07.
DR   P-PSDB; P94619.
```

```
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 1; Page 456; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   455 BP;    102 A;    107 C;    115 G;    122 T;

  Query Match             52.5%;  Score 425.2;  DB 1;  Length 455;
  Best Local Similarity   96.3%;  Pred. No. 3.8e-119;
  Matches  439;  Conservative    3;  Mismatches   12;  Indels    2;  Gaps    1;

Qy    177 GGTCTCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGC 236
          ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db      2 GGTCTCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGKCTGCTCCGGAGGCCGTGACGGC 61

Qy    237 CAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGC 296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGC 121

Qy    297 CACCTCCGCCGGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATAT 356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 CACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATAT 181

Qy    357 TTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACAC 416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 TTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACAC 241

Qy    417 AGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGA 476
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    242 AGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATNCCAGTGGCAATGA 301

Qy    477 AACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAA 536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 AACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAA 361

Qy    537 TGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGA 596
          |||||||||||||    |:      :|||||||||||||||||||||||||||||||||||
Db    362 TGGCTATTCCTAC--NNTKAGCAGTNNNWTGTCTCTTTTTCTTGGATGGTTGGGAGCAGA 419

Qy    597 TCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTT 632
          |||||||||||||||||||||||||||||||||||
Db    420 TCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTT 455


RESULT   4
T74893
ID   T74893 standard; cDNA; 1951 BP.
AC   T74893;
DT   02-OCT-1997  (first entry)
DE   Mouse neurogenic differentiation protein (NeuroD2) DNA clone 1.1.1.
KW   Neurogenic differentiation protein; NeuroD; neuroD2 gene;
KW   transcriptional activator; neuron; pancreas; gastrointestine;
KW   knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;
KW   ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             230. .1378
FT                   /*tag=  a
PN   WO9716548-A1.
PD   09-MAY-1997.
PF   30-OCT-1996; U17532.
PR   02-NOV-1995; US-552142.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA   (WEIN/) WEINTRAUB N.
PI   Hollenberg SM,  Lee JE,  Tapscott SJ,  Weintraub HM;
DR   WPI; 97-272117/24.
DR   P-PSDB; W22442.
PT   Nucleic acid encoding neurogenic differentiation polypeptide -
PT   useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT   development
PS   Claim 1; Page 69-70; 81pp; English.
CC   Neurogenic differentiation (NeuroD) genes (T74887-94) and proteins
CC   (W22436-43) from human, mouse and frog have been identified,
CC   isolated and sequence.  NeuroD polypeptides are tissue-specific
CC   basic-helix-loop-helix (bHLH) transcriptional activators involved
CC   in neuronal, endocrine and gastrointestinal development.  They were
CC   discovered by expression cloning and screening assays designed to
CC   identify possible bHLH proteins capable of interacting with the
CC   protein product of the Drosophila daugherless gene.  Novel neuroD2
CC   and neuroD3 genes, related to neuroD1, have been identified.
CC   NeuroD nucleic acids can be used to produce NeuroD polypeptides,
CC   construction of test cell lines, as probes, in gene therapy, and to
CC   produce transgenic animals as models of disease.
SQ   Sequence   1951 BP;    359 A;    668 C;    557 G;    353 T;

  Query Match              4.5%;  Score 36.4;  DB 1;  Length 1951;
  Best Local Similarity   49.0%;  Pred. No. 0.31;
  Matches   97;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180
          | | ||||   |  || |  |  ||   || |  |||   | | |||  | | | ||
Db    260 CTCCTCTCGGACGTGCCCAAGTTCGCCAGCTGGGGCGACGGCGACGACGACGAGCCGAGG 319

Qy    181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
            | | ||  | |||   ||||     | |  ||||||| | | | | | |  | ||
Db    320 AGCGACAAGGGCGACGCGCCGCCGCAGCCTTCTCCTGCTCCCGGGTCGGGGGCTCCAGGA 379

Qy    241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300
          | ||  | | |     |   || ||  |   | ||||     |    |||   ||||
Db    380 CCCGCCCGGGCCGCCAAGCCAGTGTCTCTTCGTGGAGGAGAAGAGATCCCTGAACCCACG 439

Qy    301 TCCGCCGGGGGGCGAGGAG 318
          |  || | || | |||||
Db    440 TTGGCTGAGGTCAAGGAG 457


RESULT   5
V42934
ID   V42934 standard; cDNA; 1951 BP.
AC   V42934;
DT   21-OCT-1998  (first entry)
DE   cDNA encoding murine neuroD2 protein, which is a bHLH protein.
KW   Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW   classification; medulloblastoma; mouse; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             230. .1378
FT                   /*tag=  a
FT                   /product= neuroD2
PN   US5795723-A.
PD   18-AUG-1998.
PF   07-AUG-1997; 910973.
PR   07-AUG-1997; US-910973.
PR   06-MAY-1994; US-239238.
PR   02-NOV-1995; US-552142.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI   Olson JM, Tapscott SJ;
DR   WPI; 98-466661/40.
DR   P-PSDB; W71018.
PT   Classifying neuroectodermal tumours from expression pattern of
```

```
Qy    241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300

Qy    301 TCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGT 360

Qy    361 AAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCT 420

Qy    421 CATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACA 480

Qy    481 CATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGC 540

Qy    541 TATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGA 600

Qy    601 TTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGA 660

Qy    661 ATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGA 720

Qy    721 AGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAA 780

Qy    781 ACATTTAGAAAAACGCAATTATATCCATAA 810
          ||||||||||||||||||||||||||||||
Db    781 ACATTTAGAAAAACGCAATTATATCCATAA 810
```

RESULT 2
X41191
ID X41191 standard; cDNA; 440 BP.
AC X41191;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:135.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR P-PSDB; Y12358.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 315; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 440 BP; 107 A; 103 C; 114 G; 114 T;

Query Match 53.9%; Score 436.8; DB 1; Length 440;
Best Local Similarity 99.5%; Pred. No. 1.1e-122;
Matches 436; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy    166 GAGAAAGTGTCGGTCTCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAG 225
          ||||||||||||||||||||||||||||||||||::|||||||||||||||||||||||
Db      3 GAGAAAGTGTCGGTCTCCAAGATGGCGGCCGCCTGGCSDTCTGGTCCGTCTGCTCCGGAG 62

Qy    226 GCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGG 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 GCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGG 122

Qy    286 GGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTG 345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 GGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTG 182

Qy    346 GGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGT 405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 GGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGT 242

Qy    406 ACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCC 465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 ACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCC 302

Qy    466 AGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGC 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 AGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGC 362

Qy    526 CGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGG 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 CGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGG 422

Qy    586 TTGGGAGCAGATCGATTT 603
          ||||||||||||||||||
Db    423 TTGGGAGCAGATCGATTT 440
```

RESULT 3
X41259
ID X41259 standard; cDNA; 455 BP.
AC X41259;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:203.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR P-PSDB; Y12426.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 05:23:28 ; Search time 35.22 Seconds
(without alignments)
5753.992 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

[redacted]abase : N_Geneseq_36:*

[redacted]d size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	810	100.0	810	1	X05735	Human beta-amyloid
	2	436.8	53.9	440	1	X41191	Human secreted pro
	3	425.2	52.5	455	1	X41259	Human secreted pro
	4	36.4	4.5	1951	1	T74893	Mouse neurogenic d
	5	36.4	4.5	1951	1	V42934	cDNA encoding muri
	6	35.4	4.4	110000	1	T58840_3	Continuation (4 of
c	7	34.6	4.3	4673	1	Q27189	P. yoelii SSP2 ant
	8	33.4	4.1	1624	1	N91112	Polygalacturonase
	9	33.4	4.1	1637	1	N80487	Plasmid clone pTom
	10	33.4	4.1	1636	1	Q38415	Polygalacturonase
	11	33.4	4.1	9179	1	V74411	Staphylococcus aur
c	12	33	4.1	591	1	V90286	EST clone DJ238. N
c	13	32.8	4.0	988	1	Q79079	Human flt-3 ligand
	14	32.8	4.0	988	1	V81506	Human flt3-ligand
	15	32.8	4.0	3250	1	X00726	Human aggrecan deg
	16	32.6	4.0	6524	1	T77803	cDNA encoding wild
	17	32.6	4.0	917	1	X22226	Human secreted pro
c	18	32.4	4.0	8169	1	V26609	Actinomadura hibis
	19	32.2	4.0	15672	1	Q10613	Rianodin receptor
c	20	32.2	4.0	3012	1	Q21188	Sequence encoding
	21	32.2	4.0	715	1	X20370	Borrelia burgdorfe
	22	32	4.0	110000	1	V21209_12	Continuation (13 o
	23	32	4.0	13086	1	V74327	Staphylococcus aur
	24	32	4.0	114955	1	X53491	Human adenosine A1
c	25	31.8	3.9	110000	1	X20248_01	Continuation (2 of
c	26	31.6	3.9	1075	1	Q64676	Human fibrillarin
c	27	31.6	3.9	857	1	Q79462	Human S86/S109 Flt
c	28	31.6	3.9	532	1	T67107	Helminth MIF cDNA
c	29	31.6	3.9	345	1	T67108	Helminth MIF cDNA
c	30	31.6	3.9	355	1	T67118	Helminth MIF PCR c
c	31	31.6	3.9	345	1	T67119	Helminth MIF cDNA
c	32	31.6	3.9	333	1	T67120	Helminth MIF PCR c
c	33	31.6	3.9	330	1	T67121	Helminth MIF cDNA
	34	31.4	3.9	8700	1	Q42541	BglII/HpaII fragme
c	35	31.4	3.9	8298	1	Q50168	p190 protein gene.
c	36	31.2	3.9	16079	1	N70753	Sequence of segmen
	37	31.2	3.9	24379	1	T93095	Streptomyces freno
	38	31.2	3.9	24379	1	V25925	Streptomyces roseo
c	39	31.2	3.9	1644	1	V69834	Human neural cell
c	40	31.2	3.9	16075	1	V99811	Gum gene cluster.
	41	31	3.8	2688	1	T72165	Alzheimer's diseas
	42	31	3.8	16826	1	V74357	Staphylococcus aur
c	43	31	3.8	4951	1	X13198	Enterococcus faeca
	44	30.8	3.8	5522	1	N70062	Sau3A-EcoRI 5.4 kb
	45	30.8	3.8	5212	1	N80317	Transcription cont

ALIGNMENTS

RESULT 1
X05735
ID X05735 standard; mRNA; 810 BP.
AC X05735;
DT 27-APR-1999 (first entry)
DE Human beta-amyloid peptide-binding protein (BBP) encoding mRNA.
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;
KW human; Alzheimer's disease; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1. .810
FT /*tag= a
FT /product= "BBP"
PN WO9846636-A2.
PD 22-OCT-1998.
PF 14-APR-1998; U07462.
PR 16-APR-1997; US-064583.
PA (AMHP) AMERICAN HOME PROD CORP.
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,
PI Walker SG;
DR WPI; 99-080736/07.
DR P-PSDB; W94291.
PT Polynucleotide encoding beta-amyloid peptide binding protein - used
PT to identify inhibitors of beta-amyloid peptide for treating
PT Alzheimer's disease
PS Claim 1; Pages 43-44; 59pp; English.
CC This represents a nucleotide sequence encoding a beta-amyloid peptide
CC binding protein (BBP). The polynucleotide comprising the entire BBP
CC nucleotide sequence of clone BBP1-fl is deposited under the accession
CC number ATCC 98617. The polynucleotide comprising a fragment of BBP
CC (nucleotides 202-807 of the full length BBP) of clone pEK196 is deposited
CC as ATCC 98399. Host cells transformed with a vector comprising the BBP
CC nucleic acid are used for the recombinant production of the protein. The
CC protein can be used in a method for diagnosing a disease characterised by
CC aberrant expression of human beta-amyloid protein (BAP). The protein can
CC also be used in a method for screening for compounds which regulate
CC expression of a BAP binding protein. The proteins, antibodies and
CC identified compounds can be used in the treatment or prevention of
CC Alzheimer's disease.
SQ Sequence 810 BP; 204 A; 183 C; 202 G; 221 T;

Query Match 100.0%; Score 810; DB 1; Length 810;
Best Local Similarity 100.0%; Pred. No. 2.1e-235;
Matches 810; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGCATATTTTAAAAGGGTCTCCCAATGTGATTCCACGGGCTCACGGGCAGAAGAACACG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCATATTTTAAAAGGGTCTCCCAATGTGATTCCACGGGCTCACGGGCAGAAGAACACG 60

Qy      61 CGAAGAGACGGAACTGGCCTCTATCCTATGCGAGGTCCCTTTAAGAACCTCGCCCTGTTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGAAGAGACGGAACTGGCCTCTATCCTATGCGAGGTCCCTTTAAGAACCTCGCCCTGTTG 120

Qy     121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180

Qy     181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
```

```
          re-examination
  JOURNAL   Mol. Immunol. 28 (4-5), 449-457 (1991)
  MEDLINE   91287731
COMMENT     See also X61312-4 & X65266.
FEATURES             Location/Qualifiers
     source          1. .1851
                     /organism="Carassius auratus"
                     /db_xref="taxon:7957"
                     /germline
                     /dev_stage="adult"
                     /clone_lib="GEM11 genomic"
                     /clone="6X"
     misc_feature    358. .365
                     /note="promoter octamer"
     misc_feature    633. .722
                     /note="FR1"
     misc_feature    723. .737
                     /note="CDR1"
     misc_feature    738. .779
                     /note="FR2"
     misc_feature    780. .836
                     /note="CDR2"
     misc_feature    837. .935
                     /note="FR3"
     misc_feature    937. .944
                     /note="7-mer"
     misc_feature    937. .974
                     /note="recombination signal sequence"
     misc_feature    966. .974
                     /note="9-mer"
BASE COUNT      610 a    353 c    313 g    575 t
ORIGIN


  Query Match             2.3%;  Score 19;  DB 4;  Length 1851;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     775 AATGAAACATTTAGAAAAA 793
           |||||||||||||||||||
Db    1666 AATGAAACATTTAGAAAAA 1648


RESULT  15
AF194825/c
LOCUS       AF194825     2069 bp    DNA             PLN       18-NOV-1999
DEFINITION  Bougainvillea alba NADH dehydrogenase (ndhF) gene, partial cds;
            chloroplast gene for chloroplast product.
ACCESSION   AF194825
VERSION     AF194825.1  GI:6449168
KEYWORDS    .
SOURCE      Bougainvillea alba.
  ORGANISM  Chloroplast Bougainvillea alba
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Caryophyllidae; Caryophyllales; Nyctaginaceae;
            Bougainvillea.
REFERENCE   1  (bases 1 to 2069)
  AUTHORS   Applequist,W.L. and Wallace,R.S.
  TITLE     Phylogeny of the portulacaceous cohort based on ndhF sequence data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2069)
  AUTHORS   Applequist,W.L. and Wallace,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Botany, Iowa State University, 353 Bessey
            Hall, Ames, IA 50011-1020, USA
FEATURES             Location/Qualifiers
     source          1. .2069
                     /organism="Bougainvillea alba"
                     /chloroplast
                     /db_xref="taxon:107614"
     gene            <1. .>2069
                     /gene="ndhF"
     CDS             <1. .>2069
                     /gene="ndhF"
                     /codon_start=3
                     /product="NADH dehydrogenase"
                     /protein_id="AAF08857.1"
                     /db_xref="GI:6449215"
                     /translation="LIGAGLLFFPIATKNLRRIWAFSSILVLSIVMIFSMKLSIQQIN
                     SNFIYQYVLSWTINNDFSLEFGYLIDPLTSIMSMLITTVAILVLIYSDNYMSHDRGYL
                     RFFAYMSFFNTSMLGLVISSNLIQIYIFWELVGMCSYLLIGFWFTRPIAANACQKAFV
                     TNRIGDFGLLLGILGLYWITGSFEFGDLFEIFNNLINKNEINFLFCILCTLFLFAGPV
                     AKSAQFPLHVWLPDAMEGPTPISALIHAATMVAAGIFLVARLLPLFIVIPYTMNVISF
                     IGIITVLLGATLALAQKDIKRSLAYSTMSQLGYMMLALGMGSYRTALFHLITHAYSKA
                     LLFLGSGSIIHSMEAIVGYSPDKSQNMVLMGGLARHVPITKASFLIGTLSLCGIPPLA
                     CFWSKDEILNDSWLYSPIFAILAYFTAGLTAFYMFRIYLLTFEGHLNVYLKNYSGKKS
                     NSFYSISLWGKERLPTINKNFCLLTLLIMNNKKRASFFSKTKKPHQIDGNFKKMKQSF
                     ITMTYFYNKNVYSYPHESENTMLFPLIILILFTFLIGFIGIPVNQEGIDILTKWLTPS
                     INLLHENLHNFLDWYEFVINAIVSVSIAYLGIFLAFVLYKPVYSSFKNFYLINSFDKI
                     GSKRIFWDKILYIIYNWSSNRGYIDSFYATSLIRGVRGLAELVYFFDRRIIDGIPNGF
                     GVTSFFV"
BASE COUNT      606 a    276 c    343 g    844 t
ORIGIN


  Query Match             2.3%;  Score 19;  DB 8;  Length 2069;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     672 AATTGATTTCATTCTTATT 690
           |||||||||||||||||||
Db     609 AATTGATTTCATTCTTATT 591


Search completed: February 11, 2000, 05:40:18
Job time: 904 sec
```

```
*                    gap of unknown length
*    12601    13373: contig of 773 bp in length
*                    gap of unknown length
*    13374    14175: contig of 802 bp in length
*                    gap of unknown length
*    14176    14963: contig of 788 bp in length
*                    gap of unknown length
*    14964    15761: contig of 798 bp in length
*                    gap of unknown length
*    15762    16578: contig of 817 bp in length
*                    gap of unknown length
*    16579    17389: contig of 811 bp in length
*                    gap of unknown length
*    17390    18189: contig of 800 bp in length
*                    gap of unknown length
*    18190    18975: contig of 786 bp in length
*                    gap of unknown length
*    18976    19754: contig of 779 bp in length
*                    gap of unknown length
*    19755    20513: contig of 759 bp in length
*                    gap of unknown length
*    20514    21305: contig of 792 bp in length
*                    gap of unknown length
*    21306    22099: contig of 794 bp in length
*                    gap of unknown length
*    22100    22893: contig of 794 bp in length
*                    gap of unknown length
*    22894    23680: contig of 787 bp in length
*                    gap of unknown length
*    23681    24500: contig of 820 bp in length
*                    gap of unknown length
*    24501    25273: contig of 773 bp in length
*                    gap of unknown length
*    25274    26053: contig of 780 bp in length
*                    gap of unknown length
*    26054    26808: contig of 755 bp in length
*                    gap of unknown length
*    26809    27560: contig of 752 bp in length
*                    gap of unknown length
*    27561    28379: contig of 819 bp in length
*                    gap of unknown length
*    28380    29180: contig of 801 bp in length
*                    gap of unknown length
*    29181    29976: contig of 796 bp in length
*                    gap of unknown length
*    29977    30764: contig of 788 bp in length
*                    gap of unknown length
*    30765    31567: contig of 803 bp in length
*                    gap of unknown length
*    31568    32375: contig of 808 bp in length
*                    gap of unknown length
*    32376    33153: contig of 778 bp in length
*                    gap of unknown length
*    33154    33943: contig of 790 bp in length
*                    gap of unknown length
*    33944    34715: contig of 772 bp in length
*                    gap of unknown length
*    34716    35493: contig of 778 bp in length
*                    gap of unknown length
*    35494    36284: contig of 791 bp in length
*                    gap of unknown length
*    36285    37064: contig of 780 bp in length
*                    gap of unknown length
*    37065    37875: contig of 811 bp in length
*                    gap of unknown length
*    37876    38652: contig of 777 bp in length
*                    gap of unknown length
*    38653    39420: contig of 768 bp in length
*                    gap of unknown length
*    39421    40192: contig of 772 bp in length
*                    gap of unknown length
*    40193    40973: contig of 781 bp in length
*                    gap of unknown length
*    40974    41754: contig of 781 bp in length
*                    gap of unknown length
*    41755    42558: contig of 804 bp in length
*                    gap of unknown length
*    42559    43360: contig of 802 bp in length
*                    gap of unknown length
*    43361    44170: contig of 810 bp in length
*                    gap of unknown length
*    44171    44965: contig of 795 bp in length
*                    gap of unknown length
*    44966    45756: contig of 791 bp in length
*                    gap of unknown length
*    45757    46543: contig of 787 bp in length
*                    gap of unknown length
*    46544    47348: contig of 805 bp in length
*                    gap of unknown length
*    47349    48152: contig of 804 bp in length
*                    gap of unknown length
*    48153    48951: contig of 799 bp in length
*                    gap of unknown length
*    48952    49743: contig of 792 bp in length
*                    gap of unknown length
*    49744    50547: contig of 804 bp in length
*                    gap of unknown length
*    50548    51324: contig of 777 bp in length
*                    gap of unknown length
*    51325    52121: contig of 797 bp in length
*                    gap of unknown length
*    52122    52894: contig of 773 bp in length
*                    gap of unknown length
*    52895    53677: contig of 783 bp in length
*                    gap of unknown length
*    53678    54458: contig of 781 bp in length
*                    gap of unknown length
*    54459    55225: contig of 767 bp in length
*                    gap of unknown length
*    55226    55972: contig of 747 bp in length
*                    gap of unknown length
*    55973    56789: contig of 817 bp in length
*                    gap of unknown length
*    56790    57605: contig of 816 bp in length
*                    gap of unknown length

  Query Match             2.5%;  Score 20;  DB 44;  Length 60013;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     187 ATGGCGGCCGCCTGGCCGTC 206
           ||||||||||||||||||||
Db   13672 ATGGCGGCCGCCTGGCCGTC 13653


RESULT  14
CAVHH6X/c
LOCUS       CAVHH6X      1851 bp    DNA             VRT       20-MAY-1992
DEFINITION  C.auratus immunoglobulin V-H hybridizing region (6X clone).
ACCESSION   X65266
VERSION     X65266.1  GI:62591
KEYWORDS    immunoglobulin heavy chain variable region; V-H hybridizing region;
            VH gene.
SOURCE      goldfish.
  ORGANISM  Carassius auratus
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Cyprininae; Carassius.
REFERENCE   1  (bases 1 to 1851)
  AUTHORS   Wilson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1992) M.R. Wilson, Basel Inst of Immunology,
            Grenzacherstrasse 487, Basel, SWITZERLAND
REFERENCE   2  (bases 1 to 1851)
  AUTHORS   Wilson,M.R., Middleton,D. and Warr,G.W.
  TITLE     Immunoglobulin VH genes of the goldfish, Carassius auratus: a
```

```
*                    gap of unknown length
*    39397    40222: contig of 826 bp in length
*                    gap of unknown length
*    40223    41040: contig of 818 bp in length
*                    gap of unknown length
*    41041    41854: contig of 814 bp in length
*                    gap of unknown length
*    41855    42685: contig of 831 bp in length
*                    gap of unknown length
*    42686    43496: contig of 811 bp in length
*                    gap of unknown length
*    43497    44318: contig of 822 bp in length
*                    gap of unknown length
*    44319    45133: contig of 815 bp in length
*                    gap of unknown length
*    45134    45974: contig of 841 bp in length
*                    gap of unknown length
*    45975    46759: contig of 785 bp in length
*                    gap of unknown length
*    46760    47564: contig of 805 bp in length
*                    gap of unknown length
*    47565    48379: contig of 815 bp in length
*                    gap of unknown length
*    48380    49198: contig of 819 bp in length
*                    gap of unknown length
*    49199    50012: contig of 814 bp in length
*                    gap of unknown length
*    50013    50782: contig of 770 bp in length
*                    gap of unknown length
*    50783    51575: contig of 793 bp in length
*                    gap of unknown length
*    51576    52358: contig of 783 bp in length
*                    gap of unknown length
*    52359    53161: contig of 803 bp in length
*                    gap of unknown length
*    53162    53988: contig of 827 bp in length
*                    gap of unknown length
*    53989    54797: contig of 809 bp in length
*                    gap of unknown length
*    54798    55665: contig of 868 bp in length
*                    gap of unknown length
*    55666    56479: contig of 814 bp in length
*                    gap of unknown length
*    56480    57266: contig of 787 bp in length
*                    gap of unknown length
*    57267    58073: contig of 807 bp in length
*                    gap of unknown length
*    58074    58885: contig of 812 bp in length
*                    gap of unknown length

  Query Match          2.5%;  Score 20;  DB 44;  Length 67795;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    188 TGGCGGCCGCCTGGCCGTCT 207
          ||||||||||||||||||||
Db  36064 TGGCGGCCGCCTGGCCGTCT 36083


RESULT  13
AC016814/c
LOCUS       AC016814     60013 bp    DNA             HTG       07-DEC-1999
DEFINITION  Mus musculus clone RP23-2M16, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016814
VERSION     AC016814.1  GI:6532079
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 60013)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-2M16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60013)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5141
                Center clone name: 2_M_16
            --------------.
            * NOTE: This record contains 76 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *                    gap of unknown length
            *      789     1582: contig of 794 bp in length
            *                    gap of unknown length
            *     1583     2380: contig of 798 bp in length
            *                    gap of unknown length
            *     2381     3145: contig of 765 bp in length
            *                    gap of unknown length
            *     3146     3925: contig of 780 bp in length
            *                    gap of unknown length
            *     3926     4720: contig of 795 bp in length
            *                    gap of unknown length
            *     4721     5530: contig of 810 bp in length
            *                    gap of unknown length
            *     5531     6318: contig of 788 bp in length
            *                    gap of unknown length
            *     6319     7115: contig of 797 bp in length
            *                    gap of unknown length
            *     7116     7914: contig of 799 bp in length
            *                    gap of unknown length
            *     7915     8685: contig of 771 bp in length
            *                    gap of unknown length
            *     8686     9494: contig of 809 bp in length
            *                    gap of unknown length
            *     9495    10257: contig of 763 bp in length
            *                    gap of unknown length
            *    10258    11037: contig of 780 bp in length
            *                    gap of unknown length
            *    11038    11831: contig of 794 bp in length
            *                    gap of unknown length
            *    11832    12600: contig of 769 bp in length
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-10G1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 67795)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3059
                Center clone name: 10_G_1
            --------------.
            * NOTE: This record contains 84 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      790: contig of 790 bp in length
            *                    gap of unknown length
            *      791     1589: contig of 799 bp in length
            *                    gap of unknown length
            *     1590     2411: contig of 822 bp in length
            *                    gap of unknown length
            *     2412     3206: contig of 795 bp in length
            *                    gap of unknown length
            *     3207     4010: contig of 804 bp in length
            *                    gap of unknown length
            *     4011     4799: contig of 789 bp in length
            *                    gap of unknown length
            *     4800     5607: contig of 808 bp in length
            *                    gap of unknown length
            *     5608     6417: contig of 810 bp in length
            *                    gap of unknown length
            *     6418     7207: contig of 790 bp in length
            *                    gap of unknown length
            *     7208     8023: contig of 816 bp in length
            *                    gap of unknown length
            *     8024     8833: contig of 810 bp in length
            *                    gap of unknown length
            *     8834     9641: contig of 808 bp in length
            *                    gap of unknown length
            *     9642    10443: contig of 802 bp in length
            *                    gap of unknown length
            *    10444    11240: contig of 797 bp in length
            *                    gap of unknown length
            *    11241    11974: contig of 734 bp in length
            *                    gap of unknown length
            *    11975    12772: contig of 798 bp in length
            *                    gap of unknown length
            *    12773    13577: contig of 805 bp in length
            *                    gap of unknown length
            *    13578    14393: contig of 816 bp in length
            *                    gap of unknown length
            *    14394    15136: contig of 743 bp in length
            *                    gap of unknown length
            *    15137    15949: contig of 813 bp in length
            *                    gap of unknown length
            *    15950    16770: contig of 821 bp in length
            *                    gap of unknown length
            *    16771    17577: contig of 807 bp in length
            *                    gap of unknown length
            *    17578    18351: contig of 774 bp in length
            *                    gap of unknown length
            *    18352    19153: contig of 802 bp in length
            *                    gap of unknown length
            *    19154    19956: contig of 803 bp in length
            *                    gap of unknown length
            *    19957    20766: contig of 810 bp in length
            *                    gap of unknown length
            *    20767    21588: contig of 822 bp in length
            *                    gap of unknown length
            *    21589    22405: contig of 817 bp in length
            *                    gap of unknown length
            *    22406    23225: contig of 820 bp in length
            *                    gap of unknown length
            *    23226    24052: contig of 827 bp in length
            *                    gap of unknown length
            *    24053    24873: contig of 821 bp in length
            *                    gap of unknown length
            *    24874    25686: contig of 813 bp in length
            *                    gap of unknown length
            *    25687    26497: contig of 811 bp in length
            *                    gap of unknown length
            *    26498    27313: contig of 816 bp in length
            *                    gap of unknown length
            *    27314    28113: contig of 800 bp in length
            *                    gap of unknown length
            *    28114    28895: contig of 782 bp in length
            *                    gap of unknown length
            *    28896    29709: contig of 814 bp in length
            *                    gap of unknown length
            *    29710    30526: contig of 817 bp in length
            *                    gap of unknown length
            *    30527    31339: contig of 813 bp in length
            *                    gap of unknown length
            *    31340    32149: contig of 810 bp in length
            *                    gap of unknown length
            *    32150    32955: contig of 806 bp in length
            *                    gap of unknown length
            *    32956    33770: contig of 815 bp in length
            *                    gap of unknown length
            *    33771    34582: contig of 812 bp in length
            *                    gap of unknown length
            *    34583    35406: contig of 824 bp in length
            *                    gap of unknown length
            *    35407    36226: contig of 820 bp in length
            *                    gap of unknown length
            *    36227    37023: contig of 797 bp in length
            *                    gap of unknown length
            *    37024    37804: contig of 781 bp in length
            *                    gap of unknown length
            *    37805    38579: contig of 775 bp in length
            *                    gap of unknown length
            *    38580    39396: contig of 817 bp in length
```

```
*                    gap of unknown length
*     6983     7756: contig of 774 bp in length
*                    gap of unknown length
*     7757     8556: contig of 800 bp in length
*                    gap of unknown length
*     8557     9339: contig of 783 bp in length
*                    gap of unknown length
*     9340    10124: contig of 785 bp in length
*                    gap of unknown length
*    10125    10901: contig of 777 bp in length
*                    gap of unknown length
*    10902    11678: contig of 777 bp in length
*                    gap of unknown length
*    11679    12436: contig of 758 bp in length
*                    gap of unknown length
*    12437    13215: contig of 779 bp in length
*                    gap of unknown length
*    13216    13996: contig of 781 bp in length
*                    gap of unknown length
*    13997    14785: contig of 789 bp in length
*                    gap of unknown length
*    14786    15573: contig of 788 bp in length
*                    gap of unknown length
*    15574    16356: contig of 783 bp in length
*                    gap of unknown length
*    16357    17133: contig of 777 bp in length
*                    gap of unknown length
*    17134    17913: contig of 780 bp in length
*                    gap of unknown length
*    17914    18697: contig of 784 bp in length
*                    gap of unknown length
*    18698    19486: contig of 789 bp in length
*                    gap of unknown length
*    19487    20258: contig of 772 bp in length
*                    gap of unknown length
*    20259    21029: contig of 771 bp in length
*                    gap of unknown length
*    21030    21806: contig of 777 bp in length
*                    gap of unknown length
*    21807    22578: contig of 772 bp in length
*                    gap of unknown length
*    22579    23351: contig of 773 bp in length
*                    gap of unknown length
*    23352    24109: contig of 758 bp in length
*                    gap of unknown length
*    24110    24887: contig of 778 bp in length
*                    gap of unknown length
*    24888    25666: contig of 779 bp in length
*                    gap of unknown length
*    25667    26440: contig of 774 bp in length
*                    gap of unknown length
*    26441    27215: contig of 775 bp in length
*                    gap of unknown length
*    27216    27995: contig of 780 bp in length
*                    gap of unknown length
*    27996    28790: contig of 795 bp in length
*                    gap of unknown length
*    28791    29560: contig of 770 bp in length
*                    gap of unknown length
*    29561    30341: contig of 781 bp in length
*                    gap of unknown length
*    30342    31111: contig of 770 bp in length
*                    gap of unknown length
*    31112    31892: contig of 781 bp in length
*                    gap of unknown length
*    31893    32673: contig of 781 bp in length
*                    gap of unknown length
*    32674    33452: contig of 779 bp in length
*                    gap of unknown length
*    33453    34237: contig of 785 bp in length
*                    gap of unknown length
*    34238    34997: contig of 760 bp in length
*                    gap of unknown length
*    34998    35774: contig of 777 bp in length
*                    gap of unknown length
*    35775    36560: contig of 786 bp in length
*                    gap of unknown length
*    36561    37341: contig of 781 bp in length
*                    gap of unknown length
*    37342    38101: contig of 760 bp in length
*                    gap of unknown length
*    38102    38895: contig of 794 bp in length
*                    gap of unknown length
*    38896    39674: contig of 779 bp in length
*                    gap of unknown length
*    39675    40406: contig of 732 bp in length
*                    gap of unknown length
*    40407    41186: contig of 780 bp in length
*                    gap of unknown length
*    41187    41966: contig of 780 bp in length
*                    gap of unknown length
*    41967    42696: contig of 730 bp in length
*                    gap of unknown length
*    42697    43458: contig of 762 bp in length
*                    gap of unknown length
*    43459    44247: contig of 789 bp in length
*                    gap of unknown length
*    44248    45027: contig of 780 bp in length
*                    gap of unknown length
*    45028    45813: contig of 786 bp in length
*                    gap of unknown length
*    45814    46610: contig of 797 bp in length
*                    gap of unknown length
*    46611    47377: contig of 767 bp in length
*                    gap of unknown length
*    47378    48138: contig of 761 bp in length
*                    gap of unknown length
*    48139    48920: contig of 782 bp in length
*                    gap of unknown length
*    48921    49698: contig of 778 bp in length
*                    gap of unknown length
*    49699    50482: contig of 784 bp in length
*                    gap of unknown length
*    50483    51269: contig of 787 bp in length
*                    gap of unknown length
*    51270    52043: contig of 774 bp in length
*                    gap of unknown length
*    52044    52797: contig of 754 bp in length
*                    gap of unknown length
*    52798    53596: contig of 799 bp in length
*                    gap of unknown length
*    53597    54394: contig of 798 bp in length
*                    gap of unknown length
*    54395    55172: contig of 778 bp in length
*                    gap of unknown length
*    55173    55953: contig of 781 bp in length
*                    gap of unknown length
*    55954    56730: contig of 777 bp in length
*                    gap of unknown length

  Query Match             2.5%;  Score 20;  DB 44;  Length 62216;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    187 ATGGCGGCCGCCTGGCCGTC 206
          ||||||||||||||||||||
Db  12430 ATGGCGGCCGCCTGGCCGTC 12411


RESULT  12
AC016538
LOCUS       AC016538    67795 bp    DNA             HTG       03-DEC-1999
DEFINITION  Homo sapiens clone RP11-10G1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016538
VERSION     AC016538.1  GI:6513994
KEYWORDS    HTG; HTGS_PHASE0.
```

```
                /note="5 copies of 6 mer 93 % conserved"
     repeat_region   55255. .55531
                /note="AluSq repeat: matches 1. .278 of consensus;
                incomplete repeat"
     repeat_region   55532. .55567
                /note="18 copies of 2 mer 86 % conserved"
     repeat_region   56142. .56233
                /note="MIR repeat: matches 52. .141 of consensus"
     repeat_region   57325. .57491
                /note="AluJo repeat: matches 136. .302 of consensus;
                incomplete repeat"
     repeat_region   58511. .58792
                /note="AluJo repeat: matches 300. .13 of consensus"
     repeat_region   59799. .59965
                /note="MIR repeat: matches 213. .47 of consensus"
     repeat_region   60865. .60900
                /note="9 copies of 4 mer 83 % conserved"
     repeat_region   61015. .61093
                /note="MER6 repeat: matches 866. .785 of consensus"
     repeat_region   61082. .61156
                /note="MER6 repeat: matches 75. .1 of consensus"
     repeat_region   61153. .61223
                /note="MER6 repeat: matches 797. .866 of consensus"
     repeat_region   61229. .61311
                /note="MIR repeat: matches 198. .104 of consensus"
     repeat_region   63776. .64004
                /note="MIR repeat: matches 262. .34 of consensus"
     repeat_region   64717. .64911
                /note="MIR repeat: matches 257. .68 of consensus"
     repeat_region   65581. .65898
                /note="L1MC3 repeat: matches 2156. .2483 of consensus"
     repeat_region   65834. .65919
                /note="L1MC1 repeat: matches 998. .1079 of consensus"
     repeat_region   67174. .67428
                /note="MIR repeat: matches 5. .262 of consensus"
     repeat_region   67616. .67860
                /note="MIR repeat: matches 262. .2 of consensus"
     repeat_region   67865. .68027
                /note="MIR repeat: matches 262. .82 of consensus"
     repeat_region   70575. .70610
                /note="9 copies of 4 mer 92 % conserved"
     repeat_region   71068. .71230
                /note="L1PA5 repeat: matches 730. .892 of consensus"
     repeat_region   72756. .73649
                /note="L1ME2 repeat: matches 10. .911 of consensus"
     repeat_region   73770. .73820
                /note="MER5A repeat: matches 167. .117 of consensus"
     repeat_region   74476. .74643
                /note="MIR repeat: matches 88. .262 of consensus"
     repeat_region   74962. .75050
                /note="MIR2 repeat: matches 60. .145 of consensus"
     repeat_region   77600. .77633
                /note="17 copies of 2 mer 85 % conserved"
     repeat_region   81606. .81779
                /note="MER45 repeat: matches 177. .1 of consensus"
     misc_feature    complement(<82028. .85660)
                /note="match: D16473 mRNA, Xq terminal portion"
     repeat_region   83381. .83406
                /note="13 copies of 2 mer 100 % conserved"
     repeat_region   83939. .84598
                /note="L1MC1 repeat: matches 394. .1079 of consensus"
     repeat_region   84641. .84764
                /note="MIR repeat: matches 188. .74 of consensus"

  Query Match             2.5%;  Score 20;  DB 11;  Length 98209;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    108 CCTCGCCCTGTTGCCCTTCT 127
          ||||||||||||||||||||
Db  52274 CCTCGCCCTGTTGCCCTTCT 52293
```

```
RESULT  11
AC016414/c
LOCUS       AC016414    62216 bp    DNA             HTG       26-NOV-1999
DEFINITION  Homo sapiens clone RP11-10A17, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016414
VERSION     AC016414.1  GI:6468838
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 62216)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-10A17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62216)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3030
                Center clone name: 10_A_17
            --------------.
            * NOTE: This record contains 80 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      742: contig of 742 bp in length
            *                   gap of unknown length
            *     743     1523: contig of 781 bp in length
            *                   gap of unknown length
            *    1524     2300: contig of 777 bp in length
            *                   gap of unknown length
            *    2301     3075: contig of 775 bp in length
            *                   gap of unknown length
            *    3076     3842: contig of 767 bp in length
            *                   gap of unknown length
            *    3843     4615: contig of 773 bp in length
            *                   gap of unknown length
            *    4616     5384: contig of 769 bp in length
            *                   gap of unknown length
            *    5385     6165: contig of 781 bp in length
            *                   gap of unknown length
            *    6166     6982: contig of 817 bp in length
```

```
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    772 ACTAATGAAACATTTAGAAA 791
          ||||||||||||||||||||
Db  50371 ACTAATGAAACATTTAGAAA 50390


RESULT  10
HS260J9
LOCUS       HS260J9     98209 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from PAC 260J9 on chromosome X.
ACCESSION   Z82193
VERSION     Z82193.1  GI:2370074
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1  (bases 1 to 98209)
  AUTHORS   Deadman,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1997) Chromosome X Project Group
            (http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 1997 this sequence version replaced gi:1772943.
            IMPORTANT: This sequence is not the entire insert of clone 260J9.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together
            with a note of the overlapping clone name. Note that the variations
            annotated may not be found in the sequence submission corresponding
            to
            the overlapping clone as we submit sequences with only a small
            overlap
            as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre chromosome X
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 260J9 is at 1 in this sequence. The true
            left end of clone 106C24 is at 98105.
            260J9 is from the library RPCI1 constructed at the Roswell Park
            Cancer
            Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .98209
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="RP1-260J9"
                     /clone_lib="RPCI-1"
     repeat_region   79. .201
                     /note="FLAM_A repeat: matches 1. .123 of consensus"
     repeat_region   219. .965
                     /note="L1PA9 repeat: matches 156. .907 of consensus"
     repeat_region   1367. .1705
                     /note="L1MA3 repeat: matches 1043. .686 of consensus"
     repeat_region   5449. .5470
                     /note="11 copies of 2 mer 100 % conserved"
     repeat_region   7443. .7570
                     /note="MIR repeat: matches 72. .198 of consensus"
     repeat_region   8163. .8204
                     /note="21 copies of 2 mer 88 % conserved"
     repeat_region   9000. .9244
                     /note="MIR repeat: matches 262. .12 of consensus"
     repeat_region   11270. .11393
                     /note="L1ME2 repeat: matches 433. .565 of consensus"
     repeat_region   11270. .11495
                     /note="L1MC3 repeat: matches 433. .673 of consensus"
     repeat_region   11907. .11948
                     /note="7 copies of 6 mer 81 % conserved"
     repeat_region   12400. .12522
                     /note="MIR2 repeat: matches 130. .1 of consensus"
     repeat_region   14179. .14298
                     /note="LTR9 repeat: matches 415. .298 of consensus"
     repeat_region   15048. .15170
                     /note="MIR repeat: matches 245. .123 of consensus"
     repeat_region   17438. .17750
                     /note="MER1B repeat: matches 337. .32 of consensus"
     repeat_region   18079. .18480
                     /note="L1 repeat: matches 4991. .5390 of consensus"
     repeat_region   18330. .19080
                     /note="L1PA3 repeat: matches 1. .750 of consensus"
     repeat_region   19085. .19261
                     /note="L1PA4 repeat: matches 717. .893 of consensus"
     repeat_region   19970. .20268
                     /note="AluJo repeat: matches 293. .1 of consensus"
     repeat_region   20298. .20339
                     /note="MIR repeat: matches 107. .66 of consensus"
     repeat_region   20345. .21777
                     /note="TIGGER1 repeat: matches 2418. .967 of consensus"
     repeat_region   21780. .22076
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   22088. .23029
                     /note="TIGGER1 repeat: matches 976. .24 of consensus"
     repeat_region   23342. .23367
                     /note="13 copies of 2 mer 92 % conserved"
     repeat_region   24261. .24328
                     /note="34 copies of 2 mer 81 % conserved"
     misc_feature    24779. .24954
                     /note="match: STS L42723"
     repeat_region   27874. .28303
                     /note="L1 repeat: matches 4958. .5388 of consensus"
     repeat_region   28148. .28837
                     /note="L1MA3 repeat: matches 1. .687 of consensus"
     repeat_region   28844. .29152
                     /note="AluY repeat: matches 297. .1 of consensus"
     repeat_region   29155. .29520
                     /note="L1MA4 repeat: matches 684. .1047 of consensus"
     repeat_region   33588. .33627
                     /note="20 copies of 2 mer 88 % conserved"
     repeat_region   35340. .35666
                     /note="AluY repeat: matches 301. .1 of consensus"
     repeat_region   35876. .35927
                     /note="26 copies of 2 mer 96 % conserved"
     repeat_region   35931. .36606
                     /note="MER42c repeat: matches 803. .1538 of consensus"
     repeat_region   36971. .37402
                     /note="L1MB7 repeat: matches 479. .920 of consensus"
     repeat_region   39498. .39647
                     /note="MER5B repeat: matches 154. .2 of consensus"
     repeat_region   42893. .43017
                     /note="AluSx repeat: matches 2. .127 of consensus;
                     incomplete repeat"
     repeat_region   43026. .43088
                     /note="AluSx/g repeat: matches 239. .301 of consensus;
                     incomplete repeat"
     repeat_region   44124. .44159
                     /note="18 copies of 2 mer 81 % conserved"
     repeat_region   44604. .44903
                     /note="AluSp repeat: matches 302. .2 of consensus"
     unsure          52177. .52297
                     /note="Gap closed by subcloning of CpG island fragment"
     repeat_region   52684. .52713
```

```
            *    97731    97810: gap of unknown length
            *    97811    98334: contig of 524 bp in length
            *    98335    98414: gap of unknown length
            *    98415    98985: contig of 571 bp in length
            *    98986    99065: gap of unknown length
            *    99066    99291: contig of 226 bp in length
            *    99292    99371: gap of unknown length
            *    99372    99912: contig of 541 bp in length
            *    99913    99992: gap of unknown length
            *    99993   100618: contig of 626 bp in length
            *   100619   100698: gap of unknown length
            *   100699   101398: contig of 700 bp in length

  Query Match             2.6%;  Score 21;  DB 42;  Length 113562;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    154 AGCGGAAGTGGCGAGAAAGTG 174
          |||||||||||||||||||||
     1865 AGCGGAAGTGGCGAGAAAGTG 1885


RESULT   7
AC012995/c
LOCUS       AC012995    76396 bp    DNA             HTG       03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC012995
VERSION     AC012995.1  GI:6223070
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 76396)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212198 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .76396
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    21300 a  16411 c  16836 g  21849 t
ORIGIN


  Query Match             2.6%;  Score 21;  DB 43;  Length 76396;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    138 CCTGGGCGGAGGCGGAAGCGG 158
          |||||||||||||||||||||
Db  12113 CCTGGGCGGAGGCGGAAGCGG 12093


RESULT   8
AC017975
LOCUS       AC017975    27387 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017975
VERSION     AC017975.1  GI:6553215
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 27387)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212858 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .27387
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     7801 a   5791 c   5937 g   7858 t
ORIGIN


  Query Match             2.6%;  Score 21;  DB 45;  Length 27387;
  Best Local Similarity   100.0%;  Pred. No. 0.54;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    154 AGCGGAAGTGGCGAGAAAGTG 174
          |||||||||||||||||||||
Db  19673 AGCGGAAGTGGCGAGAAAGTG 19693


RESULT   9
AP000076
LOCUS       AP000076   100000 bp    DNA             PRI       20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 8p11.2, senescence gene
            region, section 12/19, complete sequence.
ACCESSION   AP000076
VERSION     AP000076.1  GI:4579097
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isomura,M., Ikegawa,S., Kinjo,T. and Nakamura,Y.
  TITLE     DNA sequence analysis of a 1.9-Mb region on chromosome 8p11.2
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1999) to the DDBJ/EMBL/GenBank databases. Mika
            Hirakawa, Japan Science and Technology Corporation (JST), Advanced
            Database Department; 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028,
            Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     >>text_table: 0000140177.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p11.2"
     STS             43625. .43766
                     /note="SHGC-9028;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:1235120"
BASE COUNT    31453 a  18346 c  19549 g  30651 t      1 others
ORIGIN


  Query Match             2.5%;  Score 20;  DB 10;  Length 100000;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
```

```
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 93 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1      964: contig of 964 bp in length
*      965     1044: gap of unknown length
*     1045     2062: contig of 1018 bp in length
*     2063     2142: gap of unknown length
*     2143     2986: contig of 844 bp in length
*     2987     3066: gap of unknown length
*     3067     3634: contig of 568 bp in length
*     3635     3714: gap of unknown length
*     3715     4638: contig of 924 bp in length
*     4639     4718: gap of unknown length
*     4719     5291: contig of 573 bp in length
*     5292     5371: gap of unknown length
*     5372     6106: contig of 735 bp in length
*     6107     6186: gap of unknown length
*     6187     7016: contig of 830 bp in length
*     7017     7096: gap of unknown length
*     7097     8094: contig of 998 bp in length
*     8095     8174: gap of unknown length
*     8175     9007: contig of 833 bp in length
*     9008     9087: gap of unknown length
*     9088    10029: contig of 942 bp in length
*    10030    10109: gap of unknown length
*    10110    11068: contig of 959 bp in length
*    11069    11148: gap of unknown length
*    11149    11725: contig of 577 bp in length
*    11726    11805: gap of unknown length
*    11806    12395: contig of 590 bp in length
*    12396    12475: gap of unknown length
*    12476    13403: contig of 928 bp in length
*    13404    13483: gap of unknown length
*    13484    14236: contig of 753 bp in length
*    14237    14316: gap of unknown length
*    14317    15330: contig of 1014 bp in length
*    15331    15410: gap of unknown length
*    15411    16745: contig of 1335 bp in length
*    16746    16825: gap of unknown length
*    16826    17681: contig of 856 bp in length
*    17682    17761: gap of unknown length
*    17762    18569: contig of 808 bp in length
*    18570    18649: gap of unknown length
*    18650    19700: contig of 1051 bp in length
*    19701    19780: gap of unknown length
*    19781    21144: contig of 1364 bp in length
*    21145    21224: gap of unknown length
*    21225    21905: contig of 681 bp in length
*    21906    21985: gap of unknown length
*    21986    22953: contig of 968 bp in length
*    22954    23033: gap of unknown length
*    23034    24086: contig of 1053 bp in length
*    24087    24166: gap of unknown length
*    24167    25883: contig of 1717 bp in length
*    25884    25963: gap of unknown length
*    25964    27209: contig of 1246 bp in length
*    27210    27289: gap of unknown length
*    27290    28436: contig of 1147 bp in length
*    28437    28516: gap of unknown length
*    28517    29905: contig of 1389 bp in length
*    29906    29985: gap of unknown length
*    29986    31609: contig of 1624 bp in length
*    31610    31689: gap of unknown length
*    31690    32620: contig of 931 bp in length
*    32621    32700: gap of unknown length
*    32701    33457: contig of 757 bp in length
*    33458    33537: gap of unknown length
*    33538    34762: contig of 1225 bp in length
*    34763    34842: gap of unknown length
*    34843    35650: contig of 808 bp in length
*    35651    35730: gap of unknown length
*    35731    37654: contig of 1924 bp in length
*    37655    37734: gap of unknown length
*    37735    38356: contig of 622 bp in length
*    38357    38436: gap of unknown length
*    38437    40525: contig of 2089 bp in length
*    40526    40605: gap of unknown length
*    40606    42579: contig of 1974 bp in length
*    42580    42659: gap of unknown length
*    42660    44115: contig of 1456 bp in length
*    44116    44195: gap of unknown length
*    44196    45369: contig of 1174 bp in length
*    45370    45449: gap of unknown length
*    45450    47205: contig of 1756 bp in length
*    47206    47285: gap of unknown length
*    47286    49223: contig of 1938 bp in length
*    49224    49303: gap of unknown length
*    49304    51498: contig of 2195 bp in length
*    51499    51578: gap of unknown length
*    51579    52672: contig of 1094 bp in length
*    52673    52752: gap of unknown length
*    52753    54833: contig of 2081 bp in length
*    54834    54913: gap of unknown length
*    54914    56621: contig of 1708 bp in length
*    56622    56701: gap of unknown length
*    56702    58223: contig of 1522 bp in length
*    58224    58303: gap of unknown length
*    58304    59080: contig of 777 bp in length
*    59081    59160: gap of unknown length
*    59161    60156: contig of 996 bp in length
*    60157    60236: gap of unknown length
*    60237    62470: contig of 2234 bp in length
*    62471    62550: gap of unknown length
*    62551    64581: contig of 2031 bp in length
*    64582    64661: gap of unknown length
*    64662    66722: contig of 2061 bp in length
*    66723    66802: gap of unknown length
*    66803    69042: contig of 2240 bp in length
*    69043    69122: gap of unknown length
*    69123    71027: contig of 1905 bp in length
*    71028    71107: gap of unknown length
*    71108    73332: contig of 2225 bp in length
*    73333    73412: gap of unknown length
*    73413    75475: contig of 2063 bp in length
*    75476    75555: gap of unknown length
*    75556    78424: contig of 2869 bp in length
*    78425    78504: gap of unknown length
*    78505    85099: contig of 6595 bp in length
*    85100    85179: gap of unknown length
*    85180    90487: contig of 5308 bp in length
*    90488    90567: gap of unknown length
*    90568    91182: contig of 615 bp in length
*    91183    91262: gap of unknown length
*    91263    91894: contig of 632 bp in length
*    91895    91974: gap of unknown length
*    91975    92461: contig of 487 bp in length
*    92462    92541: gap of unknown length
*    92542    93207: contig of 666 bp in length
*    93208    93287: gap of unknown length
*    93288    93741: contig of 454 bp in length
*    93742    93821: gap of unknown length
*    93822    95002: contig of 1181 bp in length
*    95003    95082: gap of unknown length
*    95083    95833: contig of 751 bp in length
*    95834    95913: gap of unknown length
*    95914    96439: contig of 526 bp in length
*    96440    96519: gap of unknown length
*    96520    97094: contig of 575 bp in length
*    97095    97174: gap of unknown length
*    97175    97730: contig of 556 bp in length
```

```
RESULT   5
DMU35074
LOCUS       DMU35074     3937 bp    mRNA            INV       17-APR-1998
DEFINITION  Drosophila melanogaster centrosomal and chromosomal factor (ccf)
            mRNA, complete cds and 5' genomic segment.
ACCESSION   U35074
VERSION     U35074.1  GI:1016261
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3937)
  AUTHORS   Kodjabachian,L., Delaage,M., Maurel,C., Miassod,R., Jacq,B. and
            Rosset,R.
  TITLE     Mutations in ccf, a novel Drosophila gene encoding a chromosomal
            factor, affect progression through mitosis and interact with Pc-G
            mutations
  JOURNAL   EMBO J. 17 (4), 1063-1075 (1998)
  MEDLINE   98130597
REFERENCE   2  (bases 1 to 3937)
  AUTHORS   Kodjabachian,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-1995) Laurent Kodjabachian, Laboratoire de
            Genetique et Physiologie du Developpement, CNRS, Parc Scientifique
            de Luminy, Case 907, Marseille 13288, France
FEATURES             Location/Qualifiers
     source          1. .3937
                     /organism="Drosophila melanogaster"
                     /strain="Oregon R"
                     /db_xref="taxon:7227"
                     /map="3R; 82E"
     misc_feature    1. .425
                     /gene="corto"
                     /note="genomic DNA"
                     /db_xref="FlyBase:FBgn0010313"
     gene            1. .3937
                     /gene="corto"
                     /note="corto"
                     /allele=""
                     /db_xref="FlyBase:FBgn0010313"
     TATA_signal     293. .298
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     mutation        407
                     /gene="corto"
                     /note="at the indicated position in the A46 enhancer trap
                     line is inserted a P(lacZ, ry) transposable element."
                     /db_xref="FlyBase:FBgn0010313"
     5'UTR           426. .1478
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     mRNA            426. .3937
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     CAAT_signal     585. .590
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     TATA_signal     635. .640
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     CDS             1479. .3131
                     /gene="corto"
                     /note="required for mitotic chromosome condensation;
                     Centrosomal and Chromosomal Factor"
                     /codon_start=1
                     /product="CCF"
                     /protein_id="AAC13917.1"
                     /db_xref="GI:1016262"
                     /translation="MTMAACYANYDMSSLSHGMSALSALQQQQQQQQQQHSQTQQQHH
                     HQQQQQHMYHAAVAAHQQQLLQQQQQQQHHRHHHQPANTSSSSNSRHSHAAATQTQVA
                     AAVANSRQQQQQQQQQQQQQQQQQTASSNSNTAPAPSPQKDYSIPLHVDCSVEYELPN
                     QPKPPAGQRVEPLLMIHPCYFRKMESQRRSPFVNNMHATARAVSSSSLSSGAALGGGG
                     SGAAVATAPSSSAARRGARAATSAQQQQQQQQRYQQQQQQLRQQHQQMSQMSQQAHYP
                     QQQSSLVRQHQQQQQQQQQQRASSNQSQRQSQSQSQSQSQSHAANSAAAAAAQASIAAA
                     AAGQWDQLAALAARTALTPHHMLHPHSHYAAKGSGGEQGGGKRDAMISGSYGQTAVAS
                     GKLQQSQVQQQQPQQQQQHCLPPPPWDATSMLMDRSPMATVPSNYQAGPDTNPMRLYS
                     ATPTAGAASGGSASVGGGGAVSGAGATGAVNTATDKLSGKYRQYLRSQRMHPYAAAAS
                     LNLAAAAAAAGQTSFVPFSSAATAVAATPTFQHLPQISCYNV"
     3'UTR           3132. .3937
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     polyA_signal    3811. .3816
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
     polyA_site      3868
                     /gene="corto"
                     /db_xref="FlyBase:FBgn0010313"
BASE COUNT     1293 a   1098 c    926 g    620 t
ORIGIN


  Query Match             2.6%;  Score 21;  DB 34;  Length 3937;
  Best Local Similarity   100.0%;  Pred. No. 0.52;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     138 CCTGGGCGGAGGCGGAAGCGG 158
           |||||||||||||||||||||
Db    2117 CCTGGGCGGAGGCGGAAGCGG 2137


RESULT   6
AC011072
LOCUS       AC011072   113562 bp    DNA             HTG       20-OCT-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR48A20 (D937) RPCI-98
            48.A.20 map 55B-55D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 93 unordered pieces.
ACCESSION   AC011072
VERSION     AC011072.4  GI:6087903
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 113562)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113562)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Oct 20, 1999 this sequence version replaced gi:6067123.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
```

```
*     5970     6049: gap of unknown length
*     6050     6771: contig of 722 bp in length
*     6772     6851: gap of unknown length
*     6852     8136: contig of 1285 bp in length
*     8137     8216: gap of unknown length
*     8217     8879: contig of 663 bp in length
*     8880     8959: gap of unknown length
*     8960     9778: contig of 819 bp in length
*     9779     9858: gap of unknown length
*     9859    10614: contig of 756 bp in length
*    10615    10694: gap of unknown length
*    10695    11280: contig of 586 bp in length
*    11281    11360: gap of unknown length
*    11361    12029: contig of 669 bp in length
*    12030    12109: gap of unknown length
*    12110    12854: contig of 745 bp in length
*    12855    12934: gap of unknown length
*    12935    14181: contig of 1247 bp in length
*    14182    14261: gap of unknown length
*    14262    14972: contig of 711 bp in length
*    14973    15052: gap of unknown length
*    15053    15978: contig of 926 bp in length
*    15979    16058: gap of unknown length
*    16059    16721: contig of 663 bp in length
*    16722    16801: gap of unknown length
*    16802    17774: contig of 973 bp in length
*    17775    17854: gap of unknown length
*    17855    18669: contig of 815 bp in length
*    18670    18749: gap of unknown length
*    18750    19787: contig of 1038 bp in length
*    19788    19867: gap of unknown length
*    19868    21014: contig of 1147 bp in length
*    21015    21094: gap of unknown length
*    21095    22235: contig of 1141 bp in length
*    22236    22315: gap of unknown length
*    22316    23360: contig of 1045 bp in length
*    23361    23440: gap of unknown length
*    23441    24064: contig of 624 bp in length
*    24065    24144: gap of unknown length
*    24145    25411: contig of 1267 bp in length
*    25412    25491: gap of unknown length
*    25492    26415: contig of 924 bp in length
*    26416    26495: gap of unknown length
*    26496    27203: contig of 708 bp in length
*    27204    27283: gap of unknown length
*    27284    27908: contig of 625 bp in length
*    27909    27988: gap of unknown length
*    27989    29152: contig of 1164 bp in length
*    29153    29232: gap of unknown length
*    29233    29730: contig of 498 bp in length
*    29731    29810: gap of unknown length
*    29811    31005: contig of 1195 bp in length
*    31006    31085: gap of unknown length
*    31086    32359: contig of 1274 bp in length
*    32360    32439: gap of unknown length
*    32440    33158: contig of 719 bp in length
*    33159    33238: gap of unknown length
*    33239    33901: contig of 663 bp in length
*    33902    33981: gap of unknown length
*    33982    34630: contig of 649 bp in length
*    34631    34710: gap of unknown length
*    34711    35484: contig of 774 bp in length
*    35485    35564: gap of unknown length
*    35565    36639: contig of 1075 bp in length
*    36640    36719: gap of unknown length
*    36720    37775: contig of 1056 bp in length
*    37776    37855: gap of unknown length
*    37856    38578: contig of 723 bp in length
*    38579    38658: gap of unknown length
*    38659    39371: contig of 713 bp in length
*    39372    39451: gap of unknown length
*    39452    40368: contig of 917 bp in length
*    40369    40448: gap of unknown length
*    40449    41736: contig of 1288 bp in length
*    41737    41816: gap of unknown length
*    41817    43163: contig of 1347 bp in length
*    43164    43243: gap of unknown length
*    43244    44365: contig of 1122 bp in length
*    44366    44445: gap of unknown length
*    44446    45300: contig of 855 bp in length
*    45301    45380: gap of unknown length
*    45381    46514: contig of 1134 bp in length
*    46515    46594: gap of unknown length
*    46595    48237: contig of 1643 bp in length
*    48238    48317: gap of unknown length
*    48318    49655: contig of 1338 bp in length
*    49656    49735: gap of unknown length
*    49736    51205: contig of 1470 bp in length
*    51206    51285: gap of unknown length
*    51286    52492: contig of 1207 bp in length
*    52493    52572: gap of unknown length
*    52573    54279: contig of 1707 bp in length
*    54280    54359: gap of unknown length
*    54360    55688: contig of 1329 bp in length
*    55689    55768: gap of unknown length
*    55769    56657: contig of 889 bp in length
*    56658    56737: gap of unknown length
*    56738    58423: contig of 1686 bp in length
*    58424    58503: gap of unknown length
*    58504    60089: contig of 1586 bp in length
*    60090    60169: gap of unknown length
*    60170    61484: contig of 1315 bp in length
*    61485    61564: gap of unknown length
*    61565    62429: contig of 865 bp in length
*    62430    62509: gap of unknown length
*    62510    63286: contig of 777 bp in length
*    63287    63366: gap of unknown length
*    63367    64268: contig of 902 bp in length
*    64269    64348: gap of unknown length
*    64349    66125: contig of 1777 bp in length
*    66126    66205: gap of unknown length
*    66206    68260: contig of 2055 bp in length
*    68261    68340: gap of unknown length
*    68341    70058: contig of 1718 bp in length
*    70059    70138: gap of unknown length
*    70139    71508: contig of 1370 bp in length
*    71509    71588: gap of unknown length
*    71589    73765: contig of 2177 bp in length
*    73766    73845: gap of unknown length
*    73846    75975: contig of 2130 bp in length
*    75976    76055: gap of unknown length
*    76056    77978: contig of 1923 bp in length
*    77979    78058: gap of unknown length
*    78059    80306: contig of 2248 bp in length
*    80307    80386: gap of unknown length
*    80387    84725: contig of 4339 bp in length
*    84726    84805: gap of unknown length
*    84806    88580: contig of 3775 bp in length
*    88581    88660: gap of unknown length
*    88661    94422: contig of 5762 bp in length
*    94423    94502: gap of unknown length
*    94503    95203: contig of 701 bp in length
*    95204    95283: gap of unknown length
*    95284    95887: contig of 604 bp in length
*    95888    95967: gap of unknown length
*    95968    96613: contig of 646 bp in length

  Query Match              2.6%;  Score 21;  DB 33;  Length 158018;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     154 AGCGGAAGTGGCGAGAAAGTG 174
           |||||||||||||||||||||
Db 153844 AGCGGAAGTGGCGAGAAAGTG 153864
```

```
            *    51509    51588: gap of unknown length
            *    51589    53474: contig of 1886 bp in length
            *    53475    53554: gap of unknown length
            *    53555    55472: contig of 1918 bp in length
            *    55473    55552: gap of unknown length
            *    55553    57609: contig of 2057 bp in length
            *    57610    57689: gap of unknown length
            *    57690    60309: contig of 2620 bp in length
            *    60310    60389: gap of unknown length
            *    60390    62885: contig of 2496 bp in length
            *    62886    62965: gap of unknown length
            *    62966    65669: contig of 2704 bp in length
            *    65670    65749: gap of unknown length
            *    65750    67506: contig of 1757 bp in length
            *    67507    67586: gap of unknown length
            *    67587    71189: contig of 3603 bp in length
            *    71190    71269: gap of unknown length
            *    71270    74072: contig of 2803 bp in length
            *    74073    74152: gap of unknown length
            *    74153    79546: contig of 5394 bp in length
            *    79547    79626: gap of unknown length
            *    79627    83255: contig of 3629 bp in length
            *    83256    83335: gap of unknown length
            *    83336    87374: contig of 4039 bp in length
            *    87375    87454: gap of unknown length
            *    87455    93625: contig of 6171 bp in length
            *    93626    93705: gap of unknown length
            *    93706   101327: contig of 7622 bp in length
            *   101328   101407: gap of unknown length
            *   101408   102139: contig of 732 bp in length
            *   102140   102219: gap of unknown length
            *   102220   102861: contig of 642 bp in length
            *   102862   102941: gap of unknown length
            *   102942   103594: contig of 653 bp in length
            *   103595   103674: gap of unknown length
            *   103675   104379: contig of 705 bp in length
            *   104380   104459: gap of unknown length
            *   104460   105024: contig of 565 bp in length
            *   105025   105104: gap of unknown length
            *   105105   105611: contig of 507 bp in length
            *   105612   105691: gap of unknown length
            *   105692   106370: contig of 679 bp in length
            *   106371   106450: gap of unknown length
            *   106451   107034: contig of 584 bp in length
            *   107035   107114: gap of unknown length
            *   107115   107777: contig of 663 bp in length
            *   107778   107857: gap of unknown length
            *   107858   108516: contig of 659 bp in length
            *   108517   108596: gap of unknown length
            *   108597   109226: contig of 630 bp in length
            *   109227   109306: gap of unknown length
            *   109307   110054: contig of 748 bp in length
            *   110055   110134: gap of unknown length
            *   110135   110654: contig of 520 bp in length
            *   110655   110734: gap of unknown length
            *   110735   111413: contig of 679 bp in length
            *   111414   111493: gap of unknown length
            *   111494   112196: contig of 703 bp in length
            *   112197   112276: gap of unknown length
            *   112277   112834: contig of 558 bp in length
            *   112835   112914: gap of unknown length
            *   112915   113543: contig of 629 bp in length
            *   113544   113623: gap of unknown length
            *   113624   114241: contig of 618 bp in length
            *   114242   114321: gap of unknown length
            *   114322   115166: contig of 845 bp in length.
FEATURES             Location/Qualifiers
     source          1. .115166
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"

  Query Match             2.6%;  Score 21;  DB 33;  Length 115166;
  Best Local Similarity   100.0%;  Pred. No. 0.57;

  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     138 CCTGGGCGGAGGCGGAAGCGG 158
           |||||||||||||||||||||
Db   90770 CCTGGGCGGAGGCGGAAGCGG 90790


RESULT   4
AC008231
LOCUS       AC008231   158018 bp    DNA             HTG       02-AUG-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR12O13 (D946) RPCI-98
            12.O.13 map 57B-57D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 158 unordered pieces.
ACCESSION   AC008231
VERSION     AC008231.2  GI:5670377
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 158018)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158018)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 2, 1999 this sequence version replaced gi:5649189.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 158 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      695: contig of 695 bp in length
            *      696      775: gap of unknown length
            *      776     1457: contig of 682 bp in length
            *     1458     1537: gap of unknown length
            *     1538     2113: contig of 576 bp in length
            *     2114     2193: gap of unknown length
            *     2194     3066: contig of 873 bp in length
            *     3067     3146: gap of unknown length
            *     3147     4235: contig of 1089 bp in length
            *     4236     4315: gap of unknown length
            *     4316     5436: contig of 1121 bp in length
            *     5437     5516: gap of unknown length
            *     5517     5969: contig of 453 bp in length
```

```
     CDS             complement(join(23119. .23276,23439. .23602,23686. .23953,
                     24029. .24192,24374. .24510))
                     /gene="T23K8.6"
                     /codon_start=1

  Query Match              2.7%;  Score 22;  DB 8;  Length 53090;
  Best Local Similarity   100.0%;  Pred. No. 0.14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     620 CTTTGGGTTTGTTAAAGTTTTG 641
           |||||||||||||||||||||
Db   28604 CTTTGGGTTTGTTAAAGTTTTG 28583


RESULT   3
AC007589
LOCUS       AC007589   115166 bp    DNA             HTG       02-AUG-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR20D10 (D667) RPCI-98
            20.D.10 map 82D-82E strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 73 unordered pieces.
ACCESSION   AC007589
VERSION     AC007589.3  GI:5670563
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 115166)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 115166)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 2, 1999 this sequence version replaced gi:5649299.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 73 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1144: contig of 1144 bp in length
            *     1145     1224: gap of unknown length
            *     1225     2182: contig of 958 bp in length
            *     2183     2262: gap of unknown length
            *     2263     3020: contig of 758 bp in length
            *     3021     3100: gap of unknown length
            *     3101     3657: contig of 557 bp in length
            *     3658     3737: gap of unknown length
            *     3738     4493: contig of 756 bp in length
            *     4494     4573: gap of unknown length
            *     4574     5654: contig of 1081 bp in length
            *     5655     5734: gap of unknown length
            *     5735     6436: contig of 702 bp in length
            *     6437     6516: gap of unknown length
            *     6517     7312: contig of 796 bp in length
            *     7313     7392: gap of unknown length
            *     7393     8314: contig of 922 bp in length
            *     8315     8394: gap of unknown length
            *     8395     9075: contig of 681 bp in length
            *     9076     9155: gap of unknown length
            *     9156     9709: contig of 554 bp in length
            *     9710     9789: gap of unknown length
            *     9790    10520: contig of 731 bp in length
            *    10521    10600: gap of unknown length
            *    10601    11335: contig of 735 bp in length
            *    11336    11415: gap of unknown length
            *    11416    12536: contig of 1121 bp in length
            *    12537    12616: gap of unknown length
            *    12617    13730: contig of 1114 bp in length
            *    13731    13810: gap of unknown length
            *    13811    14642: contig of 832 bp in length
            *    14643    14722: gap of unknown length
            *    14723    16189: contig of 1467 bp in length
            *    16190    16269: gap of unknown length
            *    16270    17095: contig of 826 bp in length
            *    17096    17175: gap of unknown length
            *    17176    17975: contig of 800 bp in length
            *    17976    18055: gap of unknown length
            *    18056    18771: contig of 716 bp in length
            *    18772    18851: gap of unknown length
            *    18852    19711: contig of 860 bp in length
            *    19712    19791: gap of unknown length
            *    19792    20847: contig of 1056 bp in length
            *    20848    20927: gap of unknown length
            *    20928    22383: contig of 1456 bp in length
            *    22384    22463: gap of unknown length
            *    22464    23584: contig of 1121 bp in length
            *    23585    23664: gap of unknown length
            *    23665    25143: contig of 1479 bp in length
            *    25144    25223: gap of unknown length
            *    25224    27013: contig of 1790 bp in length
            *    27014    27093: gap of unknown length
            *    27094    28756: contig of 1663 bp in length
            *    28757    28836: gap of unknown length
            *    28837    30846: contig of 2010 bp in length
            *    30847    30926: gap of unknown length
            *    30927    32440: contig of 1514 bp in length
            *    32441    32520: gap of unknown length
            *    32521    34270: contig of 1750 bp in length
            *    34271    34350: gap of unknown length
            *    34351    36231: contig of 1881 bp in length
            *    36232    36311: gap of unknown length
            *    36312    38409: contig of 2098 bp in length
            *    38410    38489: gap of unknown length
            *    38490    40043: contig of 1554 bp in length
            *    40044    40123: gap of unknown length
            *    40124    41333: contig of 1210 bp in length
            *    41334    41413: gap of unknown length
            *    41414    43068: contig of 1655 bp in length
            *    43069    43148: gap of unknown length
            *    43149    44227: contig of 1079 bp in length
            *    44228    44307: gap of unknown length
            *    44308    45419: contig of 1112 bp in length
            *    45420    45499: gap of unknown length
            *    45500    47288: contig of 1789 bp in length
            *    47289    47368: gap of unknown length
            *    47369    49362: contig of 1994 bp in length
            *    49363    49442: gap of unknown length
            *    49443    51508: contig of 2066 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1. .53090
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="T23K8"
     gene            complement(<1. .1608)
                     /gene="T23K8.1"
     CDS             complement(join(<1. .146,426. .575,907. .974,1044. .1608))
                     /gene="T23K8.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="T23K8.1"
                     /protein_id="AAD26867.1"
                     /db_xref="GI:4646194"
                     /translation="MEEYQSNESEDKRSWIWSKAVSVGKKVLTAGVVVSSAPLLVPSL
                     FVASTLAFLSSVPFCLFLANYACTQKVMSTLLPDTEETGGVGKEDDDESGFDEYSKIG
                     HGEGAAGVGEAALFRGKEEPIPIQVKEDEEMAKESTSLLEKIRDEGRTDKETSERTLQ
                     DDKKSGNAKSEEVQEQPEKREAPETRREGETGATKIETSTGKDDEEISSNEPIDQASG
                     AQGTGEEKRKNTTKKKKKTGRAGNRFLKCHVMGYLVWRAAYPFVYSEEQLWETMETLR
                     KVVGYSVARSATCAEELKALYVFTGVVEPPRSS"
     gene            complement(2911. .7282)
                     /gene="T23K8.2"
     CDS             complement(join(2911. .3060,3150. .3239,3323. .3456,
                     3551. .3698,3779. .3985,4068. .4265,4313. .4473,4562. .4717,
                     4797. .4874,4919. .5063,5159. .5287,5364. .6806,6842. .7282))
                     /gene="T23K8.2"
                     /note="Contains similarity to F13P17.8 gi|3337355
                     hypothetical protein from Arabidopsis thaliana BAC
                     gb|AC004481.  EST gb|AA394560 comes from this gene."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD26868.1"
                     /db_xref="GI:4646195"
                     /translation="MKSSPRKLAQRYYEQGDHIKALEVIEDSILKIGGKLKLPSDLCF
                     LQGHIFMALARKTENSDLKFAFFLASVECYLEDYMLQAFAAVSLFHLGERLGSTLYYK
                     KAVKIAKRSLSVMSFDPSKSDQQKTKRTLEDIVETTGSRIIHGLSEIRIKIRSDSHTI
                     AAKRLRLYWAGMSAERKRNFMKVSTVELRSYVERVYGREGLDALEQVLDSARINRKWK
                     FWMCRTCSQTFFYPKKFKNHLEQVHDAKYKPVREDLAQSIDDVWAGMISVADWEPVDA
                     LAAAEMIKNRLEFVKEFVYVNGWSKDWPLAADEERRKLLKEIRLLLVLLLERKILSCS
                     IRDWMMRFPVQHLIAQFEVSENTITTECRLVETPQSICFLECDELNQILDLLKRIKCE
                     RDDGTELISMATDSLWRRTQVKEKIDIDHVCSLMLLDKRLLRGKIASFDDEGSIDVCD
                     HNVHYAKTNPQGDDIITWLLDDYSLIDKSFGFPRSIGAHNLDIRMAVLRAIHFTRRTL
                     ATRYAKKWQILCYDECLNDARNLCIQEDESRMNVPEDQRNIYASLLCDSCEEQLTIDT
                     EDPLFTELYLCAVRDVLEGISHPTFDFTNAEDCLELIHGLKNISDDIVLKSIDLLKSV
                     VTSKVLLADSKILLVENSRIHLLNDLIRLSVFDYRSYILPLLKRFLREELEGIVDMDA
                     KAKLAALEEELLTEEKKRKEKKSGPKKKKHRSNKVAEFVRMNVPWCSREHQQASRVIL
                     IRMTHTKRVLNMSRESSINLEPGVSLLKMVDEDSIEPEERGRQETSSNTNNHEEAIKD
                     LKNMPTKDSLSEDATRYRSALDMTLKVNQSPTKIRFKDLDSPLLIQLIRYFQTALLNI
                     KVLQEDLVHNRQPFHGNLEEQVPYALQNLFSAIVSEQIAEEGLYSYLLSNLLASLEGV
                     HSMSSDAADVVVAILEFFHCWKSPERESLVTRLFTLAEYERMSCKKCKRKPNYPEQSS
                     YGIVMAANSIRNLKCAFGNIKFEDILKMGRMKDGMICDVKTGGCGDINFVHHTISRCP
                     PIFTIVLEWVKNETEKDISETTKALDWEIDISRVYEGLEESNIKYRLVSMIGCVVEGE
                     YICMAYKKNRWVSLRHEALAEEVVGNWNNVVRFCGERKVRPEILFYEAFQWPNKWQKL
                     ISRKYQYENLAFM"
     misc_feature    9509. .9875
                     /note="Homologous to U5 small nuclear RNA gb|X13012."
     gene            complement(10409. .14859)
                     /gene="T23K8.3"
     CDS             complement(join(10409. .10498,10597. .10686,10775. .10908,
                     11005. .11152,11241. .11447,11532. .11729,11753. .11958,
                     12040. .12229,12300. .12329,12416. .12550,12633. .12761,
                     12835. .14859))
                     /gene="T23K8.3"
                     /note="Contains similarity to F13P17.8 gi|3337355
                     hypothetical protein from Arabidopsis thaliana BAC
                     gb|AC004481."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD26881.1"
                     /db_xref="GI:4646208"
                     /translation="MAPLLPLLVEDPEEAKIQKLVHGYFAEGDYIKALEVIEDSISTR
                     GKEKVPAMVHYLQGCFFYEQGTKTENSDMKFPFLLASVECFTEDCGVHAYSAIALLEL
                     GKLIGSASFYKKALNIAKEGLSFIASFGGLNLSEENTKSSLENVVLVAESMIPKLQGS
                     VRSDSDTAATEAMIEAADSMVLLQKMSEPKLRESKNIRDSIRSDLETAAGKRLRSYWA
                     SMDVESKRNFMKVSIAKLTAYVERLYGRVGLDALEQVLDSARINRKWKFWMCRSCSQK
                     FYYPKKFRSHLEQEHAAKFKPSTTKHMAQRVDEVWAGMISVAGWEPVDVVAAVEMIKT
                     QLESVKEFVYENGWSKNWPLAVDEERSKLLKELQSLLLLFQERKILSCSIRDWMMQLT
                     VKHLAQLEISEHTLSSKCRLVETPQSICFLECHELNQILDLLKRINCERVDGTELVSN
                     ATDSLCGRLRVKEKIEFDHEFSFMLLDKRLLRGKIASFDDEGIIDVCDHNVHYTKTHP
                     QGDDTITWLLDYPLIDESFEFPRSIRAHNLEICVAVLRAIHFTCATLGTKYAKKLQIL
                     NYDAALIDAKNLCILEDERRKNVPEDQWNTYASLLCDKCEERLVIDAGDSHTTKLCLC
                     AVRDVLEKASHPTFEFSDLEDCMTLIHGHKNLNDDTVLKSIDLLKSVVTNKVPVADSK
                     ILLIENSRTNLLNDLVKLSVFDYRSYILHLLKRYQRDELDEIVDMDAKTKLAAAQAEH
                     LSEEKQEKGKKSGSNKKKNRSNKRTSASMSKDDVFYEQNVFLHMSSESSVNLEPKVTP
                     PSPKSTEEDSMEPEDTLSSERGRLEISSNTKSQDETAKDMENMPGEDLLSENLESAHK
                     KVTRYNSALDMTLKVKQTPAKIRFKNLSSLLLIQLIRYFHIKSPSLLISALLNIKVLQ
                     EDLVHNRQLFHGNLEEQVPCALKDFFSAVVSEQIKEDELYSYLLSNLLASLEEVHSLS
                     SDADEVVVTILEFLHWWKSPEKESLVTRLFTLEEYERMRCTKCRRNPNYPEQSSYGIV
                     MAANSIRNLKCAFGNMKFEEILTVIRMEDEMICDVKTGGCGKTNFVHHVLSRCPPIFT
                     IVLEWEKNETETEISETAKALAWEIDISRLYEGLEEPNTNYRLVSMIGCGEEGEHICM
                     AYKTNRWVSLRHEALLEEGVGNWKNVVRICGERRVRPEILFYEAAYK"
     gene            complement(15424. .19974)
                     /gene="T23K8.4"
     CDS             complement(join(15424. .15510,15601. .15711,15775. .15905,
                     16005. .16152,16226. .16432,16538. .16726,16852. .16955,
                     17031. .17131,17513. .17596,17628. .17752,17854. .17982,
                     18064. .19974))
                     /gene="T23K8.4"
                     /note="Contains similarity to F13P17.8 gi|3337355
                     hypothetical protein from Arabidopsis thaliana BAC
                     gb|AC004481."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD26869.1"
                     /db_xref="GI:4646196"
                     /translation="MKPPSLEDCRLDESYFAGGDHIKAMVIIEDSIFKQGKGNDLLGL
                     LHFQQGSIFRAQARRTENSDVKFTFLLGSVECFAESEGFSSFTALSLFEMGQHLGSSM
                     YYKKSVEKAKEYLSVLAELDEYGPKEVKSQKDVERIIEDAESRIAGGKSLVASPIVRS
                     EPKAKESKKNPDSKLRSFWRGLLDDEFKRNFLKMEITKLISFVEAMYGKEGRDALDQVL
                     TFARENRKWRFWMCRTCSKKFSSPEECKNHLGQEHAAEFKPSSTKDMPQTVTKVWGHK
                     ISVGGWDPVDAAAAVEMIKNRLQDVKAFAYENGWSKDWPLAIDEERSKLLKEIKLLLV
                     SFWELKILSCSIRDWMMQFPVKHLAQFEVSEHTLTTECRLVETPQSICFLEFRELNQI
                     LHFLRAIKGERDDGKNLVCKAVDSFWDGTRVKEKIDFDSQFSFLLLDKRLLKCKLTRF
                     DNEGTVNVFDPNDYYANAQVHGDDILSWLADYSLGDEIFRFPKPIRTHNIDIWVAVLR
                     AVQFTCRTLRTKYEKKLRMICYDAALIDANNLCIREDERRRNIPEDQWTLYASLLCDK
                     CEENIRRHAGDSLTTKLSLCAVQDVLGGASQPTFDFTDLVDCRNLINGHKHISDDIVL
                     KSINLLKSVVTYKVPLVDSKILLVENSRISLLKDLVSLSVFDYRSYILQPVKLYLEEE
                     LEDAAPPELLSEKKQEKAKKSGSKKRNHKSTKVFDGLNSICTNEWSSCHIRELQRACL
                     VILTRMSNSKAKEDSMELEDDTKGEKSLFEISSNTNNQEEATKACHLKDMQSMPREDL
                     ETVNGKAATRYNSALDMMLKSLCNIKVLKEDLVNNRKPFSENQVPCALRDFFSAFVSE
                     KIKEEELYGHLLSNLLDSLEELHSLSSNAAEVLVAILEFCHCWKSPESESLVTRLFTL
                     EEYERMSCRKCRRKPNYPEQSSYGIVMAADSIRDLECALGNIMFEDTLKVIRMNNEMI
                     CDVGTGGCGERNLVHHTISRCPPIFTIVLEWEKNETEIEIYETTNALHWEIDISRLYE
                     GLEPNTNYRLVSMSENLGMQIGCVEEGEYICMAYEKNRWVSRRHEASAEEVVGNWKSV
                     VKICGERKIRSEVLFYEAVQ"
     gene            20566. .22178
                     /gene="T23K8.5"
     CDS             join(20560♦ .20621,20711. .20831,20885. .21034,21105. .21287,
                     21368. .21509,21593. .21729,21809. .21889,21975. .22052,
                     22131. .22178)
                     /gene="T23K8.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="T23K8.5"
                     /protein_id="AAD26870.1"
                     /db_xref="GI:4646197"
                     /translation="MEMDLSEVKEMEKDEEIDSDLSDSDTEEPEEKILDATSLSALDM
                     ILKVSQNSTIQNESVSLWNISVFVSNLSMDYVQEANQSLTDNESVPLQDILRIFMLNN
                     LSLTHPLEANGVSKLFLNILELIPRWNYHFEVNEVAKTICIRCKTDMAYSGETSYGII
                     INANSLRVYKSAFKEFTFENIIKASMINVKMLCDKEGCGKRNYVDTMISNLPSAFIVA
                     LQWENNETEKEIFDTASVLATEIDISAIYRYVGDSAFTKYRLVSMVWSHGNLYNCVAY
                     ENNRWVRHFCSEIEVIGDWDGVLRIFRELHIRPEILFFENAILKDQMFSEQGSED"
     gene            complement(23119. .24510)
                     /gene="T23K8.6"
```

```
REFERENCE   1  (bases 1 to 199723)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199723)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT      Center project name: H_NH0521M14.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2353: contig of 2353 bp in length
            *     2354     2371: gap of unknown length
            *     2372     4340: contig of 1969 bp in length
            *     4341     4358: gap of unknown length
            *     4359     6425: contig of 2067 bp in length
            *     6426     6443: gap of unknown length
            *     6444     9090: contig of 2647 bp in length
            *     9091     9108: gap of unknown length
            *     9109    12172: contig of 3064 bp in length
            *    12173    12190: gap of unknown length
            *    12191    16034: contig of 3844 bp in length
            *    16035    16052: gap of unknown length
            *    16053    19152: contig of 3100 bp in length
            *    19153    19170: gap of unknown length
            *    19171    22944: contig of 3774 bp in length
            *    22945    22962: gap of unknown length
            *    22963    27695: contig of 4733 bp in length
            *    27696    27713: gap of unknown length
            *    27714    31855: contig of 4142 bp in length
            *    31856    31873: gap of unknown length
            *    31874    35780: contig of 3907 bp in length
            *    35781    35798: gap of unknown length
            *    35799    40716: contig of 4918 bp in length
            *    40717    40734: gap of unknown length
            *    40735    45236: contig of 4502 bp in length
            *    45237    45254: gap of unknown length
            *    45255    48959: contig of 3705 bp in length
            *    48960    48977: gap of unknown length
            *    48978    55337: contig of 6360 bp in length
            *    55338    55355: gap of unknown length
            *    55356    62799: contig of 7444 bp in length
            *    62800    62817: gap of unknown length
            *    62818    70545: contig of 7728 bp in length
            *    70546    70563: gap of unknown length
            *    70564    77953: contig of 7390 bp in length
            *    77954    77971: gap of unknown length
            *    77972    84023: contig of 6052 bp in length
            *    84024    84041: gap of unknown length
            *    84042    90480: contig of 6439 bp in length
            *    90481    90498: gap of unknown length
            *    90499    97338: contig of 6840 bp in length
            *    97339    97356: gap of unknown length
            *    97357   108436: contig of 11080 bp in length
            *   108437   108454: gap of unknown length
            *   108455   116696: contig of 8242 bp in length
            *   116697   116714: gap of unknown length
            *   116715   124591: contig of 7877 bp in length
            *   124592   124609: gap of unknown length
            *   124610   135616: contig of 11007 bp in length
            *   135617   135634: gap of unknown length
            *   135635   147194: contig of 11560 bp in length
            *   147195   147211: gap of unknown length
            *   147212   159038: contig of 11827 bp in length
            *   159039   159055: gap of unknown length
            *   159056   178182: contig of 19127 bp in length
            *   178183   178199: gap of unknown length
            *   178200   199723: contig of 21524 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199723
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-521M14"
BASE COUNT    57984 a  40505 c  40199 g  60534 t    501 others
ORIGIN


  Query Match             2.8%;  Score 23;  DB 44;  Length 199723;
  Best Local Similarity   100.0%;  Pred. No. 0.037;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     606 CCTTGGATACCCTGCTTTGGGTT 628
           |||||||||||||||||||||||
Db  102200 CCTTGGATACCCTGCTTTGGGTT 102222


RESULT   2
T23K8/c
LOCUS       T23K8       53090 bp    DNA             PLN       13-MAY-1999
DEFINITION  Arabidopsis thaliana chromosome 1 BAC T23K8 sequence, complete
            sequence.
ACCESSION   AC007230
VERSION     AC007230.2  GI:4582411
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 53090)
  AUTHORS   Vysotskaia,V.S., Schwartz,J.R., Yu,G., Toriumi,M., Lenz,C., Liu,S.,
            Li,J., Kremenetskaia,I., Ngan,I., Luros,J., Gonzalez,A., Altafi,H.,
            Araujo,R., Chao,Q., Conn,L., Conway,A.B., Dunn,P., Hansen,N.,
            Huizar,L., Kim,C., Palm,C.J., Rowley,D., Shinn,P., Walker,M.,
            Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
  TITLE     Arabidopsis thaliana chromosome 1 BAC T23K8 sequence
  JOURNAL   Unpublished (1999)
  REMARK    This sequence is of BAC T23K8 from Arabidopsis thaliana chromosome
            1.  The sequence does not represent the  sequence of the entire
            insert of this clone.  It is shorter by 31564 bp because we submit
            only the unique sequence of the clone.  However, in order to
            facilitate the joining of overlapping clones in the future for
            creation of larger contigs, we provide a small overlap (200 bp)
            between overlapping submitted clones.  The 3' end of this sequence
            overlaps by 200 bp the 5' end of the sequence of the BAC T8F5.
REFERENCE   2  (bases 1 to 53090)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   3  (bases 1 to 53090)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   4  (bases 1 to 53090)
  AUTHORS   Theologis.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1999) Plant Gene Expression Center, 800 Buchanan
            St., Albany, CA 94710, USA
REFERENCE   5  (bases 1 to 53090)
  AUTHORS   Theologis.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) Plant Gene Expression Center, 800 Buchanan
            St., Albany, CA 94710, USA
COMMENT     On Apr 13, 1999 this sequence version replaced gi:4567183.
            The sequence of BAC T23K8 from Arabidopsis thaliana chromosome 1.
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 05:25:14 ; Search time 553.5 Seconds
(without alignments)
-4443.489 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: OLIGO_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold : 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	2.8	199723	44	AC016756	AC016756 Homo sapi
c 2	22	2.7	53090	8	T23K8	AC007230 Arabidops
3	21	2.6	115166	33	AC007589	AC007589 Drosophil
4	21	2.6	158018	33	AC008231	AC008231 Drosophil
5	21	2.6	3937	34	DMU35074	U35074 Drosophila
6	21	2.6	113562	42	AC011072	AC011072 Drosophil
c 7	21	2.6	76396	43	AC012995	AC012995 Drosophil
8	21	2.6	27387	45	AC017975	AC017975 Drosophil
9	20	2.5	100000	10	AP000076	AP000076 Homo sapi
10	20	2.5	98209	11	HS260J9	Z82193 Human DNA s
c 11	20	2.5	62216	44	AC016414	AC016414 Homo sapi
12	20	2.5	67795	44	AC016538	AC016538 Homo sapi
c 13	20	2.5	60013	44	AC016814	AC016814 Mus muscu
c 14	19	2.3	1851	4	CAVHH6X	X65266 C.auratus i
c 15	19	2.3	2069	8	AF194825	AF194825 Bougainvi
16	19	2.3	89137	8	ATAC006284	AC006284 Arabidops
c 17	19	2.3	194430	10	CNS01DS2	AL121652 BAC seque
c 18	19	2.3	171556	10	CNS01DS8	AL121658 BAC seque
c 19	19	2.3	220895	10	HS250D10	Z99716 Human DNA s
c 20	19	2.3	2124	10	HSM801125	AL117592 Homo sapi
c 21	19	2.3	126963	11	HS753D4	AL031676 Human DNA
22	19	2.3	16437	14	AF123770	AF123770 Shuttle v
c 23	19	2.3	18772	14	AF133437	AF133437 Cloning v
c 24	19	2.3	16297	14	CVU75991	U75991 Cloning vec
c 25	19	2.3	19500	14	CVU75992	U75992 Cloning vec
c 26	19	2.3	11612	14	CVU80929	U80929 Cloning vec
c 27	19	2.3	8279	19	HSA131497	Aj131497 Homo Sapi
28	19	2.3	53517	32	AC002405	AC002405 Homo sapi
c 29	19	2.3	197010	32	AL133159	AL133159 Mus muscu
c 30	19	2.3	197010	32	AL133160	AL133160 Mus muscu
c 31	19	2.3	151245	32	AL133230	AL133230 Homo sapi
32	19	2.3	151210	32	AP000457	AP000457 Homo sapi
33	19	2.3	105079	32	AP000632	AP000632 Homo sapi
c 34	19	2.3	176875	32	CNS0000K	AL049840 Homo sapi
c 35	19	2.3	206273	32	CNS0000S	AL049870 Homo sapi
c 36	19	2.3	190880	32	CNS00M8N	AL079307 Homo sapi
c 37	19	2.3	177617	32	CNS00M8T	AL079343 Homo sapi
c 38	19	2.3	216550	32	CNS00M8U	AL079351 Homo sapi
39	19	2.3	240250	32	CNS00YVG	AL096821 Homo sapi
c 40	19	2.3	183059	32	CNS01DRP	AL117692 Homo sapi
c 41	19	2.3	201658	32	CNS01DRX	AL121579 Homo sapi
c 42	19	2.3	243962	32	CNS01DRY	AL121594 Homo sapi
c 43	19	2.3	216986	32	CNS01DS1	AL121612 Homo sapi
c 44	19	2.3	194957	32	CNS01DTI	AL132777 Homo sapi
45	19	2.3	176667	45	AC016927	AC016927 Homo sapi

ALIGNMENTS

RESULT 1
AC016756
LOCUS AC016756 199723 bp DNA HTG 04-DEC-1999
DEFINITION Homo sapiens clone RP11-521M14, *** SEQUENCING IN PROGRESS ***, 29 unordered pieces.
ACCESSION AC016756
VERSION AC016756.1 GI:6524288
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
  Query Match             2.1%;  Score 17;  DB 1;  Length 2625;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     335 ACCTCAAAGTGGGACAA 351
           |||||||||||||||||
Db    1223 ACCTCAAAGTGGGACAA 1239


RESULT  15
T31984
ID   T31984 standard; DNA; 8075 BP.
AC   T31984;
DT   16-JAN-1997  (first entry)
DE   Invertebrate calcium channel subunit coding sequence.
KW   Invertebrate calcium channel subunit; pharmacology; receptor;
KW   screening; identification; drug; insecticide;
     Drosophila melanogaster; ss.
     Drosophila melanogaster.
     Key             Location/Qualifiers
FT   cds             157. .7708
FT                   /*tag= a
FT                   /product= Calcium channel subunit
PN   WO9622307-A1.
PD   25-JUL-1996.
PF   19-JAN-1996; U00887.
PR   19-JAN-1995; US-374888.
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
PI   Dubald MM,  Hall LM,  Ren D,  Zheng W;
DR   WPI; 96-354475/35.
DR   P-PSDB; W01884.
PT   Gene encoding invertebrate calcium channel subunit - useful for
PT   screening cpds. for activity as pesticides and drugs
PS   Claim 7; Page 70-80; 149pp; English.
CC   The gene encoding the neuronal invertebrate calcium channel
CC   subunit can be used to transform host cells which then express
CC   functional calcium channels with the same pharmacological properties
CC   as receptors in neuronal invertebrate tissue.  Cells carrying the
CC   calcium channel can be used to identify insecticides (especially
CC   those having no harmful effects on vertebrates) and drugs.
SQ   Sequence   8075 BP;    2093 A;    1935 C;    2081 G;    1966 T;

  Query Match             2.1%;  Score 17;  DB 1;  Length 8075;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

     148 GGCGGAAGCGGAAGTGG 164
         |||||||||||||||||
Db  7432 GGCGGAAGCGGAAGTGG 7448


Search completed: February 11, 2000, 03:37:01
Job time: 211 sec
```

```
OS   Acremonium chrysogenum.
FH   Key             Location/Qualifiers
FT   cds             1252. .2717
FT                   /*tag= a
FT                   /product= PGK
FT   intron          1318. .1462
FT                   /*tag= b
FT   intron          1884. .1947
FT                   /*tag= c
PN   J05213998-A.
PD   24-AUG-1993.
PF   02-AUG-1991; 282369.
PR   03-AUG-1990; JP-204978.
PR   30-JUL-1991; JP-189984.
PA   (ASAH ) ASAHI CHEM IND CO LTD.
DR   WPI; 93-299652/38.
DR   P-PSDB; R49247.
PT   Novel polypeptide obtd. by culturing transformed fungus - having
PT   blood coagulation preventing, platelet aggregation preventing and
PT   thrombolytic activities
PS   Disclosure; Fig 61; 65pp; Japanese.
CC   PGK is described in Reference Example 2 (1).
CC   Novel polypeptides, obtd. by culturing transformed fungus, have
CC   blood coagulation preventing, platelet aggregation preventing
CC   and thrombolytic activities.
CC   In an example, plasmid M13mp19TMD3 (constructed from pSV2TMJ2
CC   (ATCC 67238) contg. a human thrombomodulin sequence) was subjected
CC   to site directed mutagenesis to prepare pSV2TMD7.  Plasmid pSV2TMD7
CC   was transfected to COS-1 cells.  The activity of promoting protein C
CC   activation by thrombin of the peptide produced by the transformed
CC   COS-1 cell was measured.  The amt. of the peptide was determined.
SQ   Sequence   3306 BP;    787 A;    1055 C;    885 G;    579 T;


  Query Match             2.1%;  Score 17;  DB 1;  Length 3306;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    126 CTCCCTCCCGCTCCTGG 142
          |||||||||||||||||
Db    209 CTCCCTCCCGCTCCTGG 193


RESULT  13
Q73500
ID   Q73500 standard; DNA; 8438 BP.
AC   Q73500;
DT   15-MAY-1995  (first entry)
DE   DNA encoding Pseudorabies virus large latency transcript.
KW   Pseudorabies virus; PRV; LLV; large latency transcript;
KW   attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICP0;
KW   protecting animals; deletion mutants; swine; ds.
OS   Pseudorabies virus.
FH   Key             Location/Qualifiers
FT   misc_feature    1. .7013
FT                   /*tag= a
FT                   /note= "derived from PRV strain InFh"
FT   misc_feature    7014. .8425
FT                   /*tag= b
FT                   /note= "derived from PRV strain Ka"
FT   cds             622. .6498
FT                   /*tag= c
FT                   /note= "encodes predicted amino acid sequence of ORF2"
FT   tata_signal     1. .6
FT                   /*tag= d
FT   misc_feature    34
FT                   /*tag= e
FT                   /note= "RNA cap site"
FT   polya_signal    8382. .8387
FT                   /*tag= f
PN   US5352596-A.
PD   04-OCT-1994.
PF   11-SEP-1992; 945283.
PR   11-SEP-1992; US-945283.
PA   (USDA ) US SEC OF AGRIC.
PI   Cheung AK,  Wesley RD;
DR   WPI; 94-316187/39.
DR   P-PSDB; R60620.
PT   New pseudorabies virus mutants for use in vaccine - having a
PT   deletion and/or insertion in the early protein O gene or large
PT   latency transcript gene
PS   Disclosure; Column 15-30; 43pp; English.
CC   Q73500 shows the Pseudorabies virus (PRV) large latency transcript
CC   (LLT). The basic sequence is derived from PRV strain InFh and PRV
CC   strain Ka. The LLT overlaps and is transcribed in the opposite
CC   orientation with respect to the EP0 (early polypeptide 0) and the
CC   immediately early gene (IE180). EP0 is nonessential for replicatio,
CC   LLT is the only gene expressed during PRV latency, and the IE180
CC   gene is absolutely necessary for PRV replication. However there are
CC   2 copies of IE180 in the genome. It is expected that PRV lacking one
CC   of the IE180 copies is viable. Deletions in the non-overlapping
CC   regions of these 3 genes will generate single deletion routants,
CC   while deletions in overlapping regions will generate double deletion
CC   mutants. The invention is concerned with the construction of attenuated
CC   viruses which have a reduced ability to reactivate from latency. This
CC   can be achieved by functionally disabling the expression of the EP0
CC   gene, or by disrupting the synthesis of the LLT, or both.(See also
CC   Q73501 and R60620-24)
SQ   Sequence   8438 BP;    1141 A;    2916 C;    3327 G;    1054 T;


  Query Match             2.1%;  Score 17;  DB 1;  Length 8438;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    182 CCAAGATGGCGGCCGCC 198
          |||||||||||||||||
Db   7222 CCAAGATGGCGGCCGCC 7238


RESULT  14
Q87970
ID   Q87970 standard; cDNA; 2625 BP.
AC   Q87970;
DT   06-NOV-1995  (first entry)
DE   Peptidyl C-terminal alpha-amidating enzyme cDNA.
KW   Peptidyl C-terminal alpha-amidating enzyme; AE; cell culture;
KW   trichostatin; CHO; ds.
OS   Not specified.
FH   Key             Location/Qualifiers
FT   cds             1. .2625
FT                   /*tag= a
FT   signal_peptide  1. .117
FT                   /*tag= b
FT   mat_peptide     118. .2625
FT                   /*tag= c
PN   EP-649900-A.
PD   26-APR-1995.
PF   07-SEP-1994; 306587.
PR   08-SEP-1993; JP-257881.
PA   (FURU/) FURUKAWA K.
PA   (SUNR ) SUNTORY LTD.
PI   Furukawa K,  Ohsuye K,  Sugimura K;
DR   WPI; 95-156754/21.
DR   P-PSDB; R73053.
PT   Increasing protein prodn. from cultured animal cells - by adding
PT   a trichostatin to the medium, effective at low concn. and not
PT   injurious to host cells
PS   Disclosure; Page 10-15; 19pp; English.
CC   3mu-1S cells (i.e. CHO cells expressing a peptidyl C-terminal alpha-
CC   amidase enzyme (AE)) were suspended in F-12 medium to which aliquots
CC   of trichostatin were added.  Cells were cultured for 3 days at 37 deg
CC   and then assayed for AE.  Without trichostatin, AE productivity was
CC   708 U/ml.  At 10, 50, 100 and 200 nM trichostatin, AE productivities
CC   were respectively 866, 1897, 1894 and 3359 U/ml.
SQ   Sequence   2625 BP;    761 A;    563 C;    633 G;    668 T;
```

```
DR   WPI; 92-010570/02.
DR   P-PSDB; R20112.
PT   Novel DNA encoding peptidyl hydroxy:glycine N-C lyase (PHL) -
PT   used to prepare PHL which can be used in the amidation of
PT   peptide(s) e.g. human calcitonin.
PS   Claim 9; Page 18; 28pp; English.
CC   The sequence represents an insert from pAE-III-202-4 (FERM BP-3172).
CC   The plasmid was prepd. by ligation of fragments from pAE-II-202-1
CC   and pAE-III-202-2 (see feature table), two subclones of AE-III-202,
CC   a clone isolated from a cDNA library in lambda gt10. The library
CC   as prepd. from RNA extracted from the skin of two heads of Xenopus
CC   laevis.  The vector serves as a source for a DNA fragment encoding
CC   PHL for the construction of an expression vector for the prepn. of
CC   recombinant PHL.  The PHL catalyses the reaction: R-GlyOH -> R-NH2.
CC   It can be used to produce peptides with amidated C-termini, e.g.
CC   calcitonin, growth hormone, LH-RH.
SQ   Sequence   3383 BP;    970 A;    702 C;    795 G;    916 T;

  Query Match             2.1%;  Score 17;  DB 1;  Length 3383;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    335 ACCTCAAAGTGGGACAA 351
          |||||||||||||||||
Db   1241 ACCTCAAAGTGGGACAA 1257


RESULT  10
Q22202/c
ID   Q22202 standard; DNA; 3306 BP.
AC   Q22202;
DT   03-JUL-1992  (first entry)
DE   A. chrysogenum phosphoglycerate kinase.
KW   Fungal PGK; yeast; ss.
OS   Acremonium chrysogenum.
FH   Key             Location/Qualifiers
FT   exon            1252. .1317
FT                   /*tag=  a
FT                   /number= 1
FT   intron          1318. .1462
FT                   /*tag=  b
FT                   /number= 1
FT   exon            1463. .1883
FT                   /*tag=  c
FT                   /number= 2
FT   intron          1884. .1947
FT                   /*tag=  d
FT                   /number= 2
FT   exon            1948. .2717
FT                   /*tag=  e
FT                   /number= 3
PN   EP-474273-A.
PD   11-MAR-1992.
PF   05-AUG-1991; 202009.
PR   03-AUG-1990; JP-204978.
PA   (ASAH ) ASAHI KASEI KOGYO.
PI   Zushi M, Gomi K, Yamamoto S, Suzuki K, Matsuda A;
DR   WPI; 92-081820/11.
DR   P-PSDB; R22025.
PT   New polypeptide inhibits blood coagulation and platelet
PT   aggregation - promotes thrombin catalysed protein C activation
PT   for treating myocardial infarction, thrombosis, embolism, etc.
PS   Example 2(Ref); Fig 61; 112pp; English.
CC   The phosphoglycerate kinase gene was isolated from a total DNA
CC   of A.chrysogenum cloned in a Lambda library. Its isolation and
CC   sequencing are described in Reference example 2-(1).
SQ   Sequence   3306 BP;    787 A;    1042 C;    898 G;    579 T;

  Query Match             2.1%;  Score 17;  DB 1;  Length 3306;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    126 CTCCCTCCCGCTCCTGG 142
          |||||||||||||||||
Db    209 CTCCCTCCCGCTCCTGG 193


RESULT  11
Q23005/c
ID   Q23005 standard; DNA; 3306 BP.
AC   Q23005;
DT   31-JUL-1992  (first entry)
DE   Phosphoglycerate kinase gene (incl. promoter).
KW   PGK; fermentation; cephalosporin; promoter; ss.
OS   Acremonium chrysogenum.
FH   Key             Location/Qualifiers
FT   promoter        1. .1251
FT                   /*tag=  a
FT                   /note= "claim 1, page 1"
FT   cds             1252. .2717
FT                   /*tag=  b
FT                   /product= PGK
FT   exon            1252. .1317
FT                   /*tag=  c
FT                   /number= 1
FT   intron          1318. .1462
FT                   /*tag=  d
FT                   /number= 1
FT   exon            1463. .1883
FT                   /*tag=  e
FT                   /number= 2
FT   intron          1884. .1947
FT                   /*tag=  f
FT                   /number= 2
FT   exon            1948. .2114
FT                   /*tag=  g
FT                   /number= 3
PN   J04058891-A.
PD   25-FEB-1992.
PF   27-JUN-1990; 166566.
PR   27-JUN-1990; JP-166566.
PA   (ASAH ) ASAHI CHEMICAL IND KK.
DR   WPI; 92-111675/14.
DR   P-PSDB; R22095.
PT   Acremonium chrysogenum promoter for efficient gene expression -
PT   comprises DNA fragment having promoter active part of
PT   phospho-glycerate kinase gene of Acremonium chrysogenum, etc.
PS   Disclosure; Fig 4; 18pp; Japanese.
CC   Various genes can be expressed efficiently in A. chrysogenum
CC   using the promoter indicated in the features.  The fermentation
CC   yield of cephalosporin can be improved using vectors contg. the
CC   promoter.
SQ   Sequence   3306 BP;    787 A;    1041 C;    899 G;    579 T;

  Query Match             2.1%;  Score 17;  DB 1;  Length 3306;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    126 CTCCCTCCCGCTCCTGG 142
          |||||||||||||||||
Db    209 CTCCCTCCCGCTCCTGG 193


RESULT  12
Q48534/c
ID   Q48534 standard; DNA; 3306 BP.
AC   Q48534;
DT   30-MAR-1994  (first entry)
DE   PGK.
KW   Transformation; fungus; blood coagulation; prevention; platelet; PGK;
KW   aggregation; thrombolytic activity; thrombolysis; human thrombomodulin;
KW   site-directed mutagenesis; promotion; protein C; activation;
KW   thrombin; ss.
```

```
ID   N92871 standard; DNA; 1665 BP.
AC   N92871;
DT   27-JUN-1990  (first entry)
DE   Expression plasmid pUCPlCI799  RV.
KW   alpha-amidating; pAX799; alpha amide; ds.
OS   Synthetic.
PN   EP-299790-A.
PD   18-JAN-1989.
PF   15-JUL-1988; 306508.
PR   17-JUL-1987; JP-177184.
PR   05-DEC-1987; JP-306867.
PA   (SUNR) Suntory Ltd.
PI   Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
DR   WPI; 89-017279/03.
DR   P-PSDB; P94857.
PT   Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
PT   and their precursors deoxyribonucleic acid encoding sequences.
PS   Disclosure; 7pp; English.
CC   The sequence encodes a derivative of the mature C-terminal alpha-
CC   amidating enzyme from plasmid pXA799.
CC   The plasmid was screened from an E.coli library using plasmid pXA457
CC   to screen a larger library.
CC   Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC   hydrophobic elements suggesting a membrane function.
CC   See also N93060.
SQ   Sequence   1665 BP;   480 A;   367 C;   393 G;   425 T;


  Query Match             2.1%;  Score 17;  DB 1;  Length 1665;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    335 ACCTCAAAGTGGGACAA 351
          |||||||||||||||||
Db   1106 ACCTCAAAGTGGGACAA 1122


RESULT   7
N92872
ID   N92872 standard; DNA; 1518 BP.
AC   N92872;
DT   27-JUN-1990  (first entry)
DE   Expression plasmid pUCPlCI799  SalI.
KW   alpha-amidating; pAX799; alpha amide; ds.
OS   Synthetic.
PN   EP-299790-A.
PD   18-JAN-1989.
PF   15-JUL-1988; 306508.
PR   17-JUL-1987; JP-177184.
PR   05-DEC-1987; JP-306867.
PA   (SUNR) Suntory Ltd.
PI   Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
DR   WPI; 89-017279/03.
DR   P-PSDB; P94858.
PT   Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
PT   and their precursors deoxyribonucleic acid encoding sequences.
PS   Disclosure; 7pp; English.
CC   The sequence encodes a derivative of the mature C-terminal alpha-
CC   amidating enzyme from plasmid pXA799.
CC   The plasmid was screened from an E.coli library using plasmid pXA457
CC   to screen a larger library.
CC   Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC   hydrophobic elements suggesting a membrane function.
CC   See also N93060.
SQ   Sequence   1518 BP;   445 A;   321 C;   361 G;   391 T;


  Query Match             2.1%;  Score 17;  DB 1;  Length 1518;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    335 ACCTCAAAGTGGGACAA 351
          |||||||||||||||||
Db   1106 ACCTCAAAGTGGGACAA 1122


RESULT   8
N94527
ID   N94527 standard; DNA; 2079 BP.
AC   N94527;
DT   27-JUN-1990  (first entry)
DE   Expression plasmid pUCPlCI799  BglII.
KW   alpha-amidating; pAX799; alpha amide; ds.
OS   Synthetic.
PN   EP-299790-A.
PD   18-JAN-1989.
PF   15-JUL-1988; 306508.
PR   17-JUL-1987; JP-177184.
PR   05-DEC-1987; JP-306867.
PA   (SUNR) Suntory Ltd.
PI   Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
DR   WPI; 89-017279/03.
DR   P-PSDB; P94856.
PT   Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
PT   and their precursors deoxyribonucleic acid encoding sequences.
PS   Disclosure; 7pp; English.
CC   The sequence encodes a derivative of the mature C-terminal alpha-
CC   amidating enzyme from plasmid pXA799.
CC   The plasmid was screened from an E.coli library using plasmid pXA457
CC   to screen a larger library.
CC   Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC   hydrophobic elements suggesting a membrane function.
CC   See also N93060.
SQ   Sequence   2079 BP;   606 A;   445 C;   500 G;   528 T;


  Query Match             2.1%;  Score 17;  DB 1;  Length 2079;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    335 ACCTCAAAGTGGGACAA 351
          |||||||||||||||||
Db   1106 ACCTCAAAGTGGGACAA 1122


RESULT   9
Q20269
ID   Q20269 standard; cDNA; 3383 BP.
AC   Q20269;
DT   06-APR-1992  (first entry)
DE   Gene encoding AE-III (peptidylhydroxyglycine N-C lyase precursor).
KW   Amidation; PHL; ds.
OS   Xenopus laevis.
FH   Key             Location/Qualifiers
FT   cds             31. .2838
FT                   /*tag=  a
FT                   /product= AE-III
FT                   /note= "including PAM and PHL domains"
FT   mat_peptide     1177. .2145
FT                   /*tag=  b
FT                   /product= PHL
FT   misc_feature    1. .1891
FT                   /*tag=  c
FT                   /note=" from pAE-III-202-1"
FT   misc_feature    1891. .3383
FT                   /*tag=  d
FT                   /note=" from pAE-III-202-2"
PN   EP-465404-A.
PD   08-JAN-1992.
PF   27-MAY-1991; 810399.
PR   01-JUN-1990; JP-141678.
PR   10-AUG-1990; JP-210535.
PR   30-NOV-1990; JP-329911.
PA   (CIBA ) CIBA GEIGY AG.
PI   Iwasaki Y, Shimoi H, Suzuki K, Ghisalba D, Nishikawa Y;
PI   Kawahara T, Kangawa K;
```

```
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   455 BP;    102 A;    107 C;    115 G;    122 T;


  Query Match            35.1%;  Score 284;  DB 1;  Length 455;
  Best Local Similarity  99.7%;  Pred. No. 7.2e-138;
  Matches  334;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    215 CTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTA 274
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     40 CTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTA 99

Qy    275 CAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGG 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100 CAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGG 159

Qy    335 ACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAAC 394
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 ACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAAC 219

Qy    395 CAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTT 454
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 CAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTT 279

Qy    455 GTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGC 514
        |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 GTAAGGATNCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGC 339

Qy    515 CCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
        |||||||||||||||||||||||||||||||||||
Db    340 CCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 374


RESULT   4
X40660/c
ID   X40660 standard; cDNA; 386 BP.
AC   X40660;
DT   18-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID No: 260.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9906550-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1232.
PR   01-AUG-1997; US-905144.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153780/13.
DR   P-PSDB; Y11938.
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
PS   Claim 1; Page 410; 675pp; English.
CC   X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins expressed in prostate, and encode the proteins given in
CC   Y11716 to Y11993 respectively. The proteins given represent the signal
CC   peptide and an N-terminal fragment of a secreted protein. The nucleic
CC   acid sequences can be used for producing secreted human gene products.
CC   They can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell proliferation and
CC   differentiation activity, haematopoiesis regulating activity, tissue
CC   growth regulating activity, reproductive hormone regulating activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC   activity or other activities. The products can be used in forensic, gene
CC   therapy and chromosome mapping procedures. The sequences can also be used
CC   for obtaining corresponding promoter sequences. The nucleic acids
CC   encoding the signal peptides can be used for directing extracellular
CC   secretion of a polypeptide or the insertion of a polypeptide into a
CC   membrane, or importing a polypeptide into a cell.
SQ   Sequence   386 BP;    79 A;    122 C;    111 G;    70 T;


  Query Match             2.2%;  Score 18;  DB 1;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    275 CAGGACCCTGGGGGGCTG 292
        ||||||||||||||||||
Db    215 CAGGACCCTGGGGGGCTG 198


RESULT   5
N90791
ID   N90791 standard; DNA; 3315 BP.
AC   N90791;
DT   10-MAR-1993  (revised)
DT   27-JUN-1990  (first entry)
DE   cDNA of plasmid pXA799 encoding C-terminal prepro-C-terminal alpha-
DE   amidating enzyme.
KW   alpha-amidating; pAX799; alpha amide; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             157. .2721
FT                   /*tag= a
PN   EP-299790-A.
PD   18-JAN-1989.
PF   15-JUL-1988; 306508.
PR   17-JUL-1987; JP-177184.
PR   05-DEC-1987; JP-306867.
PA   (SUNR) Suntory Ltd.
PI   Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
DR   WPI; 89-017279/03.
DR   P-PSDB; P94854.
PT   Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
PT   and their precursors deoxyribonucleic acid encoding sequences.
PS   Disclosure; 7pp; English.
CC   Plasmid pXA799 contains a sequence derived from Xenopus laevis.
CC   The plasmid was screened from an E.coli library using plasmid pXA457
CC   to screen a larger library.
CC   Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC   hydrophobic elements suggesting a membrane function.
CC   See also N93060.
SQ   Sequence   3315 BP;    940 A;    702 C;    763 G;    910 T;


  Query Match             2.1%;  Score 17;  DB 1;  Length 3315;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    335 ACCTCAAAGTGGGACAA 351
        |||||||||||||||||
Db   1379 ACCTCAAAGTGGGACAA 1395


RESULT   6
N92871
```

```
Qy    241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACC 300

Qy    301 TCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGT 360

Qy    361 AAAGATCCAAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGATCCAAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCT 420

Qy    421 CATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACA 480

Qy    481 CATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGC 540

Qy    541 TATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGA 600

Qy    601 TTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGA 660

Qy    661 ATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGA 720

Qy    721 AGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAA 780

Qy    781 ACATTTAGAAAAACGCAATTATATCCATAA 810
          |||||||||||||||||||||||||||||
Db    781 ACATTTAGAAAAACGCAATTATATCCATAA 810
```

RESULT 2
X41191
ID X41191 standard; cDNA; 440 BP.
AC X41191;
DT 17-JUN-1999 (first entry)
Human secreted protein 5' EST SEQ ID NO:135.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR P-PSDB; Y12358.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 315; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 440 BP; 107 A; 103 C; 114 G; 114 T;

Query Match 49.3%; Score 399; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    205 TCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTC 264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     42 TCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTC 101

Qy    265 TCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTT 324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    102 TCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTT 161

Qy    325 AAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCT 384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    162 AAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCT 221

Qy    385 ACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCC 444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    222 ACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCC 281

Qy    445 AACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGT 504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    282 AACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGT 341

Qy    505 TTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCA 564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    342 TTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCA 401

Qy    565 TTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTT 603
          |||||||||||||||||||||||||||||||||||||||
Db    402 TTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTT 440
```

RESULT 3
X41259
ID X41259 standard; cDNA; 455 BP.
AC X41259;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:203.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR P-PSDB; Y12426.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 456; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 03:33:30 ; Search time 36.32 Seconds
(without alignments)
5579.724 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: OLIGO_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	810	100.0	810	1	X05735	Human beta-amyloid
	2	399	49.3	440	1	X41191	Human secreted pro
	3	284	35.1	455	1	X41259	Human secreted pro
c	4	18	2.2	386	1	X40660	Human secreted pro
	5	17	2.1	3315	1	N90791	cDNA of plasmid pX
	6	17	2.1	1665	1	N92871	Expression plasmid
	7	17	2.1	1518	1	N92872	Expression plasmid
	8	17	2.1	2079	1	N94527	Expression plasmid
	9	17	2.1	3383	1	Q20269	Gene encoding AE-I
c	10	17	2.1	3306	1	Q22202	A. chrysogenum pho
c	11	17	2.1	3306	1	Q23005	Phosphoglycerate k
c	12	17	2.1	3306	1	Q48534	PGK. Novel polypep
	13	17	2.1	8438	1	Q73500	DNA encoding Pseud
	14	17	2.1	2625	1	Q87970	Peptidyl C-termina
	15	17	2.1	8075	1	T31984	Invertebrate calci
	16	17	2.1	8075	1	T31981	Neuronal invertebr
c	17	17	2.1	20199	1	V52139	Streptococcus pneu
	18	17	2.1	6669	1	V55039	Human HIAP-1 codin
	19	17	2.1	318	1	V88049	EST clone FQ505. N
	20	17	2.1	23439	1	V74349	Staphylococcus aur
	21	17	2.1	11307	1	X12986	Enterococcus faeca
c	22	17	2.1	872	1	X14223	H. pylori GHPO 122
	23	17	2.1	481	1	V99731	Human adult testis
c	24	17	2.1	110000	1	X20248_08	Continuation (9 of
	25	16	2.0	3640	1	N82026	Cloned p2.1 insert
	26	16	2.0	1729	1	N81256	EcoRI-BamHI fragme
	27	16	2.0	3573	1	N81065	Sequence of human
	28	16	2.0	3373	1	N80892	Sequence encoding
	29	16	2.0	3440	1	Q13967	Bce-4 gene with re
	30	16	2.0	4175	1	Q14099	C-ski protein gene
	31	16	2.0	4055	1	Q14799	FB29 chicken c-ski
	32	16	2.0	1729	1	N40059	Sequence encoding
	33	16	2.0	30	1	Q26869	Primer 1851-M; mut
c	34	16	2.0	1470	1	Q54828	IL-2 receptor gamm
	35	16	2.0	26	1	Q72741	Interleukin-2R gam
c	36	16	2.0	2542	1	Q87430	Human cystathionin
	37	16	2.0	2870	1	Q91177	Lipase and accesso
	38	16	2.0	966	1	Q93034	Wild tomato endoch
	39	16	2.0	935	1	Q93035	Wild tomato endoch
c	40	16	2.0	10897	1	T09187	MuTu putative onco
	41	16	2.0	5233	1	T58225	DNA encoding inosi
c	42	16	2.0	80073	1	T58840_5	Continuation (6 of
	43	16	2.0	3680	1	T59684	Gene locus for 2,4
	44	16	2.0	5076	1	T59685	Genes for 2,4-diac
	45	16	2.0	6387	1	T59686	Genes for (modulat

ALIGNMENTS

RESULT 1
X05735
ID X05735 standard; mRNA; 810 BP.
AC X05735;
DT 27-APR-1999 (first entry)
DE Human beta-amyloid peptide-binding protein (BBP) encoding mRNA.
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;
KW human; Alzheimer's disease; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1. .810
FT /*tag= a
FT /product= "BBP"
PN WO9846636-A2.
PD 22-OCT-1998.
PF 14-APR-1998; U07462.
PR 16-APR-1997; US-064583.
PA (AMHP) AMERICAN HOME PROD CORP.
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,
PI Walker SG;
DR WPI; 99-080736/07.
DR P-PSDB; W94291.
PT Polynucleotide encoding beta-amyloid peptide binding protein - used
PT to identify inhibitors of beta-amyloid peptide for treating
PT Alzheimer's disease
PS Claim 1; Pages 43-44; 59pp; English.
CC This represents a nucleotide sequence encoding a beta-amyloid peptide
CC binding protein (BBP). The polynucleotide comprising the entire BBP
CC nucleotide sequence of clone BBP1-fl is deposited under the accession
CC number ATCC 98617. The polynucleotide comprising a fragment of BBP
CC (nucleotides 202-807 of the full length BBP) of clone pEK196 is deposited
CC as ATCC 98399. Host cells transformed with a vector comprising the BBP
CC nucleic acid are used for the recombinant production of the protein. The
CC protein can be used in a method for diagnosing a disease characterised by
CC aberrant expression of human beta-amyloid protein (BAP). The protein can
CC also be used in a method for screening for compounds which regulate
CC expression of a BAP binding protein. The proteins, antibodies and
CC identified compounds can be used in the treatment or prevention of
CC Alzheimer's disease.
SQ Sequence 810 BP; 204 A; 183 C; 202 G; 221 T;

Query Match 100.0%; Score 810; DB 1; Length 810;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 810; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGCATATTTTAAAAGGGTCTCCCAATGTGATTCCACGGGCTCACGGGCAGAAGAACACG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCATATTTTAAAAGGGTCTCCCAATGTGATTCCACGGGCTCACGGGCAGAAGAACACG 60

Qy      61 CGAAGAGACGGAACTGGCCTCTATCCTATGCGAGGTCCCTTTAAGAACCTCGCCCTGTTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGAAGAGACGGAACTGGCCTCTATCCTATGCGAGGTCCCTTTAAGAACCTCGCCCTGTTG 120

Qy     121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGGAAGTGGCGAGAAAGTGTCGGTC 180

Qy     181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCCAAGATGGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGA 240
```

```
  Best Local Similarity   22.0%;  Pred. No. 20;
  Matches   37;  Conservative   19;  Mismatches   53;  Indels   59;  Gaps    7;

Qy      72 SAPEAVTARLVGVLWFVSV-----------TTGPWGAVATSAGGEESLKCEDLKVGQYI 119
           | |:   |:: | ::  ||           | | |       |  |:: | :: |  :
Db     455 SCPQNKAAKVAGYVYLTSVGGLVHGTIQIKATAGYWFT------GGNSVQ-ESIRFGLVL 507

Qy     120 CKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDS------SGNETHFTGNEVGFFKPI 173
           |          |:  : ||::| :||| :   ||      : |   :| | |
Db     508 C----------PFSARDPTANLSGWPAPVVWSGDSNTPLYFAANAISYTNNRVNL---- 553

Qy     174 SCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGI 221
             | |  ||   |              |      ||  |  |:
Db     553 ---AVTGNFYKEETEL---------------PGYTRHSFCPTGTTGM 581


RESULT  15
O90297
ID   O90297      PRELIMINARY;      PRT;  3432 AA.
AC   O90297;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GP78 POLYPROTEIN.
OS   Japanese encephalitis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GP78;
RA   VRATI S., GIRI R.K., RAZDAN A., MALIK P., AGARWAL V., WANI S.A.;
RT   "Molecular characterization of an Indian isolate of Japanese
RT   encephalitis virus that shows extended lag phase during growth.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF075723; AAC27708.1; -.
DR   HSSP; P14336; 1SVB.
DR   PFAM; PF00869; Flavi_glycoprot; 1.
DR   PFAM; PF00949; Flavi_helicase; 1.
DR   PFAM; PF01004; Flavi_M; 1.
DR   PFAM; PF00948; Flavi_NS1; 1.
DR   PFAM; PF01005; Flavi_NS2A; 1.
DR   PFAM; PF01002; Flavi_NS2B; 1.
DR   PFAM; PF01350; Flavi_NS4A; 1.
DR   PFAM; PF01349; Flavi_NS4B; 1.
DR   PFAM; PF00972; Flavi_NS5; 1.
DR   PFAM; PF01570; Flavi_propep; 1.
DR   PFAM; PF01003; Flavi_capsid; 1.
KW   Polyprotein.
SQ   SEQUENCE   3432 AA;  379788 MW;  9C74D083 CRC32;


  Query Match             5.7%;  Score 81.5;  DB 12;  Length 3432;
  Best Local Similarity   22.8%;  Pred. No. 1.5e+02;
  Matches   71;  Conservative   31;  Mismatches   93;  Indels  117;  Gaps   19;

Qy      47 LGGG--GSGSGE--------KVSVSKM----AAAWPSGPSAP--------------EAV 77
           || |    | ||          |  : |:    |  |   |  |              :
Db    2787 LGSGTRAVGKGEIHSNQEKIKKRIQKLKEEFATTWHKDPEHPYRTWTYHGSYEVKATGSA 2846

Qy      78 TARLVGVLWFVSVTTGPWGAVA---------TSAGGEESLKCEDLKVGQYICKDPKINDA 128
           :: : ||   | : : || |:|          |:  |:: :  |  |||    : |
Db    2847 SSLVNGV---VKLMSKPWDAIANVTTMAMTDTTPFGQQRVFKE--KVGTMAPEPPA---G 2898

Qy     129 TQEPVN-CTNYT-AHVSCFPAPNITCKD------------------------------ 155
            :| :|  ||:  ||:|    | :  |:
Db    2899 AKEVLNETTNWLWAHLSREKRPRLCTKEEFIKKVNSNAALGAVFAEQNQWSTAREAVDDP 2958

Qy     155 ------SSGNETHFTG-------NEVG--FFKPISCRNVNGYSYKVAVALSLFLGWLGAD 199
                 | |  |       | :|    ||       |   |   : :::  ||||
Db    2959 RFWEMVDEERENHLRGECHTCIYNMMGKREKKP-------GEFGKAKGSRAIWFMWLGAR 3011

Qy     200 RFYLGYPALGLL--------KFCTVGFCGIG-SLIDFILISMQIVGPSDGSSYIIDYYG- 250
             || : ||| |        :    |  | |   : :||    | |   |  |  |  |
Db    3012 --YLEFEALGFLNEDHWLSRENSGGGVEGSGVQKLGYIL--RDIAGKQGGKMYADDTAGW 3067

Qy     250 -TRLTRLSITNE 260
            ||:||  : ||
Db    3068 DTRITRTDLENE 3079


Search completed: February 11, 2000, 04:33:42
Job time: 26 sec
```

```
DR   PFAM; PF00394; Cu-oxidase; 1.
KW   Oxidoreductase.
FT   NON_TER       1      1
SQ   SEQUENCE   409 AA;  44056 MW;  0219A2E5 CRC32;


  Query Match           5.7%;  Score 82;  DB 10;  Length 409;
  Best Local Similarity   22.6%;  Pred. No. 10;
  Matches   47;  Conservative   24;  Mismatches   73;  Indels   64;  Gaps    8;

Qy      63 MAAAWP--SGPSAPEAVTAR-------LVGVLWFVSVTTGPWGAVATSAGGEESLKCEDL 113
           | || |    | | : :||         :|       :| |  | : :    ::||   |
Db     132 MVAASPFMDAPIAVDNITATATLHYSGALGTSPTTLTSTPPQNATSVANNFLDALK--SL 189

Qy     114 KVGQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPI 173
             :|  | |:  |      :   :|| :   | |  ||| ::|:    : | | |  |
Db     190 NSKKYPAKVPQTVD------HSLFFTAGLGINPCP--TCKQANGSRVVASVNNVTFVMP- 241

Qy     174 SCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQ 233
                     ||:  : | |  |    |   :||
Db     241 ----------TVALLQAHFFGINGV--FTTDFPA------------------------- 263

Qy     234 IVGPSDGSSYIIDYYGTRLTRLSITNET 261
                 :: :| ||  | |: || |
Db     263 ------NPPFVFNYTGTPPTNLATTNGT 284


RESULT  12
O88354
ID   O88354      PRELIMINARY;      PRT;   465 AA.
AC   O88354;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HEART PANCREATIC LIPASE.
GN   PL-H.
OS   Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Spermophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE; 98318668.
RA   ANDREWS M.T., SQUIRE T.L., BOWEN C.M., ROLLINS M.B.;
RT   "Low-temperature carbon utilization is regulated by novel gene
RT   activity in the heart of a hibernating mammal.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:8392-8397(1998).
DR   EMBL; AF027293; AAC40162.1; -.
DR   HSSP; P29183; 1HPL.
DR   PROSITE; PS00120; LIPASE_SER; 1.
DR   PFAM; PF00151; lipase; 1.
DR   PFAM; PF01477; PLAT; 1.
DR   PRINTS; PR00821; TAGLIPASE.
DR   PRINTS; PR00823; PANCLIPASE.
SQ   SEQUENCE   465 AA;  51246 MW;  477C9CB1 CRC32;


  Query Match           5.7%;  Score 82;  DB 11;  Length 465;
  Best Local Similarity   21.6%;  Pred. No. 12;
  Matches   45;  Conservative   31;  Mismatches   84;  Indels   48;  Gaps   10;

Qy      14 AHGQKNTRRDG-----TGLYPMRGPFKNL---------------ALLPFSLPLLGGGGSG 53
           | |:   | :|     |||| |    |:                 |:     |::   | |
Db     174 AAGEAGRRTNGAIGRITGLDPAEPCFEGTPELVRLDPSDAQFVDAIHTDGAPIVPNLGFG 233

Qy      54 SGEKVSVSKMAAAWPSG----PSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLK 109
             : |        :|:|    |   : : :::| :        | |       |  :
Db     234 MSQTVGHLDF---FPNGGIEMPGCQKNILSQIVDI-------DGIW------EGTRDFAA 277

Qy     110 CEDLKVGQYICKDPKINDATQEPVNCTNYTAHVS--CFPAPNITCKDSSGNETHFTG--N 165
           |  |:  :|   | :|       :| :|:   :  ||| |:  |          ::|  |
Db     278 CNHLRSYKYY-TDSIVNPTGFAAFSCASYSVFSANKCFPCPSGGCPQMGHYADRYSGKTN 336

Qy     166 EVGFFKPISCRNVNGYS---YKVAVALS 190
             ||    ::  : | :|   |||:| ||
Db     337 GVGQKFYLNTGDKNNFSRLRYKVSVTLS 364


RESULT  13
P87656
ID   P87656      PRELIMINARY;      PRT;   644 AA.
AC   P87656;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   FIBER.
OS   Eggdrop syndrome-1976 virus.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae;
OC   unclassified Adenoviridae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=127;
RA   AKOPIAN T.A., DORONIN K.K., ZAKHARCHUK A.N., LOBANOV V.A.,
RA   NARODITSKY B.S.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z86065; CAB06667.1; -.
DR   PFAM; PF00608; adeno_fiber2; 4.
SQ   SEQUENCE   644 AA;  67647 MW;  5A7D86D3 CRC32;


  Query Match           5.7%;  Score 81.5;  DB 12;  Length 644;
  Best Local Similarity   22.0%;  Pred. No. 20;
  Matches   37;  Conservative   19;  Mismatches   53;  Indels   59;  Gaps    7;

Qy      72 SAPEAVTARLVGVLWFVSV-----------TTGPWGAVATSAGGEESLKCEDLKVGQYI 119
           | |:   |:: | ::  ||            | | |        |  |:: | :: |  :
Db     455 SCPQNKAAKVAGYVYLTSVGGLVHGTIQIKATAGYWFT------GGNSVQ-ESIRFGLVL 507

Qy     120 CKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDS------SGNETHFTGNEVGFFKPI 173
           |           | :  : ||::| :|||  :   ||      : |   :| | |
Db     508 C----------PFSARDPTANLSGWPAPVVWSGDSNTPLYFAANAISYTNNRVNL---- 553

Qy     174 SCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGI 221
               | |  ||    |                |      ||  |  |:
Db     553 ---AVTGNFYKEETEL---------------PGYTRHSFCPTGTTGM 581


RESULT  14
O11424
ID   O11424      PRELIMINARY;      PRT;   644 AA.
AC   O11424;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   FIBER PROTEIN.
OS   avian adenovirus EDS.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=127;
RX   MEDLINE; 98042451.
RA   HESS M., BLOECKER H., BRANDT P.;
RT   "The complete nucleotide sequence of the egg drop syndrome virus: an
RT   intermediate between mastadenoviruses and aviadenoviruses.";
RL   Virology 238:145-156(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=127;
RA   HESS M.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y09598; CAA70814.1; -.
DR   PFAM; PF00608; adeno_fiber2; 4.
SQ   SEQUENCE   644 AA;  67677 MW;  E150260D CRC32;


  Query Match           5.7%;  Score 81.5;  DB 12;  Length 644;
```

```
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL034443; CAA22367.1; -.
SQ   SEQUENCE   446 AA;  47069 MW;  712D79A5 CRC32;


  Query Match             5.8%;  Score 84;  DB 2;  Length 446;
  Best Local Similarity   19.7%;  Pred. No. 7.7;
  Matches   52;  Conservative   34;  Mismatches  110;  Indels   68;  Gaps    9;

Qy      39 LLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLWF--VSVTTGPWG 96
           |:      :|| | | |    || :   :|  |       |  : :: |   ::   ||
Db     106 LIVLGYGFVGGIGLGIGYISPVSTLIKWFPDRPG-----MATGIAIMGFGGGALIASPWS 160

Qy      97 AVATSAGGEES-----------LKCEDLKVGQYICKDPK-----------------IN 126
           |    : | ::              : :|  : : |:
Db     161 AQMLKSFGTDNSGIALAFLVHGLTYAVFMLLGVLLVRVPRPRERADGRPAPLEGVQVSAR 220

Qy     127 DATQEP-------VNCTNYTAHVSCF--PAPNITCKDSSGNETHFTGNEVGFFKPISCRN 177
            | : |       | | | || :      || ||   |  :         ||   :|  |
Db     221 SAVRTPQFWLLWIVLCMNVTAGIGILEKAAPMITDFFSDTSTPVSVTAAAGFVALLSAAN 280

Qy     178 VNGYSYKVAVALSLFLGW------LGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILIS 231
           : |           ||      :|    |  | :| | :  :   |  |   |:| :
Db     281 MAGR-----------FGWSSASDLIGRKNIYRVYLGVGALMYTLIALFGDSSKPLFVLCA 329

Qy     232 MQIV-----GPSDGSSYIIDYYGT 250
           : :|     | :   :|: | :||
Db     330 LVVVSFYGGGFATAPAYLKDLFGT 353


RESULT   9
P93125
ID   P93125      PRELIMINARY;      PRT;  1274 AA.
AC   P93125;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   TTF-1.
GN   TTF1.
OS   Dunaliella salina.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Dunaliellaceae; Dunaliella.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FISHER M., GOKHMAN I., PICK U., ZAMIR A.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U77059; AAB36531.1; -.
DR   HSSP; P56410; 1AOV.
DR   PFAM; PF00405; transferrin; 6.
SQ   SEQUENCE   1274 AA;  136667 MW;  1CB5A0D3 CRC32;


  Query Match             5.8%;  Score 83;  DB 10;  Length 1274;
  Best Local Similarity   22.9%;  Pred. No. 33;
  Matches   30;  Conservative   17;  Mismatches   50;  Indels   34;  Gaps    6;

Qy      57 KVSVSKMAAAWPSGPSAPEAV-TARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKV 115
           :|    :   |     || :    |:|         | :| |  : ||
Db     579 QVDAETIEKFWEDNVCAPGSTENGPLIG--------GGKYGEVGENGGG---------- 620

Qy     116 GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISC 175
              :||  | :  :::|     |   | |         |  ||:  ||| :: | |  :
Db     620 ---LCKRCKTDCTSEDPY--AGYDGAVHCI-------DDDDGNQ--FTGGDIAFVKHSTL 665

Qy     176 RNVNGYSYKVA 186
           |: || :   |
Db     666 RDYNGPNLNTA 676


RESULT  10
Q9XBP8
ID   Q9XBP8      PRELIMINARY;      PRT;   789 AA.
AC   Q9XBP8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SERINE/THREONINE KINASE PKN11.
GN   PKN11.
OS   Myxococcus xanthus.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DZF1;
RA   INOUYE S., JAIN R., UEKI T., NARIYA H., XU C., HSU M.,
RA   MUNOZ-DORADO J., FAREZ-VIDAL E., INOUYE M.;
RT   "Sequence Analysis of 13 Eukaryotic-like Protein Ser/Thr Kinases of
RT   Myxococcus xanthus, a Developmental Bacterium and Significance of
RT   Their Coexistence with Protein His Kinases.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159690; AAD42854.1; -.
KW   Kinase.
SQ   SEQUENCE   789 AA;  81613 MW;  FB7C3265 CRC32;


  Query Match             5.7%;  Score 82.5;  DB 2;  Length 789;
  Best Local Similarity   26.6%;  Pred. No. 21;
  Matches   49;  Conservative   16;  Mismatches   76;  Indels   43;  Gaps    8;

Qy       5 KGSP-NVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLP---------------FSLPLLG 48
           :|:| |  | ||         |    |  |   : |  |               |   :||
Db     500 EGTPSNTEPEAHPGMRAAEPGASEVPAAGAHASGAEGPAASARASGAEGRAHAFGSGVLG 559

Qy      49 GGGSGSGEKV--------SVSKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTG---PWGA 97
              ||| :|         : |  | |  ||| |  || |         ::|  |     |
Db     560 TSARGSGAEVPAGAHASGAESLEAGARASGPVASVAVPAARRQWAMGLAVALGLAVVGGG 619

Qy      98 VATSAGGEESLKCEDLKVGQYICKDPKI-NDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
           ||   | ||:          :  :| :   |  |||     ||     |||:   :::
Db     620 VAVMRGTEEA-------APPAVPAEPMVAAPAVDEPVP----TAQ----PAPSAVSENAE 664

Qy     157 GNET 160
           | ||
Db     665 GGET 668


RESULT  11
P93368
ID   P93368      PRELIMINARY;      PRT;   409 AA.
AC   P93368;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   DIPHENOL OXIDASE (EC 1.10.3.2) (LACCASE) (URISHIOL OXIDASE)
DE   (FRAGMENT).
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC   Nicotiana.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97080572.
RA   KIEFER-MEYER M., GOMORD V., O'CONNELL A., HALPIN C., FAYE L.;
RT   "Cloning and sequence analysis of laccase-encoding cDNA clones from
RT   tobacco.";
RL   Gene 178:205-207(1996).
CC   -!- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE + 2
CC       H(2)O.
CC   -!- COFACTOR: COPPER.
DR   EMBL; U45243; AAC49538.1; -.
DR   MENDEL; 9402; Nicta;1457;9402.
```

```
RA   WEISS S.R.;
RL   Virology 0:0-0(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHV-A59;
RA   WEISS S.R., LEPARC-GOFFART I., HINGLEY S.T.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF029248; AAB86819.1; -.
DR   PFAM; PF01601; Corona_S2; 2.
SQ   SEQUENCE   1324 AA;  145899 MW;  EF34B05C CRC32;


  Query Match             6.1%;  Score 88;  DB 12;  Length 1324;
  Best Local Similarity   23.6%;  Pred. No. 12;
  Matches   59;  Conservative   26;  Mismatches   95;  Indels   70;  Gaps   13;

Qy      25 TGLYPMRG-PFKNLALLP--------FSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSA-- 74
           || ||: |  |:|||           | | |     |   |  |  :  : ||| :|
Db      66 TGYYPVDGSKFRNLALTGTNSVSLSWFQPPYLSQFNDGIFAK--VQNLKTSTPSGATAYF 123

Qy      74 PEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQY-ICKDPKINDATQEP 132
           |  |   | |   : :|   |:  |  ::            | || ||  |
Db     124 PTIVIGSLFGYTSY-TVVIEPYNGVIMAS------------VCQYTICLLP--------- 162

Qy     133 VNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE-VGFF----KPISCRNVNGYSYKVAV 187
             :  | |  |              ||: :||:    ||  |     ::  |
Db     162 --------YTDCKPNTN-------------GNKLIGFWHTDVKPPICVLKRNFTLNVNA 199

Qy     188 ALSLFLGWLGADRFYLGY---PALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
             |  :     ||  |   |:     |       || ::    :   |  |: ||::
Db     200 DAFYFHFYQHGGTFYAYYADKPSATTFLFSVY----IGDILTQYYVLPFICNPTAGSTFA 255

Qy     245 IDYYGTRLTR 254
           |: | | :
Db     256 PRYWVTPLVK 265


RESULT   6
Q21535
ID   Q21535      PRELIMINARY;      PRT;  1207 AA.
AC   Q21535; Q22233;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   T05C12.10 PROTEIN.
GN   T05C12.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   THOMAS K.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z49968; CAA90265.1; -.
DR   EMBL; Z66500; CAA90265.1; JOINED.
DR   EMBL; Z66500; CAA91313.1; -.
DR   EMBL; Z49968; CAA91313.1; JOINED.
DR   HSSP; P06620; 1INA.
DR   PFAM; PF01079; Hint; 1.
SQ   SEQUENCE   1207 AA;  121407 MW;  52AFD3CE CRC32;


  Query Match             6.1%;  Score 87.5;  DB 5;  Length 1207;
  Best Local Similarity   21.1%;  Pred. No. 12;
  Matches   47;  Conservative   26;  Mismatches   83;  Indels   67;  Gaps    8;

Qy       2 HILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSVS 61
           |  : | : ||    : ||   |  : |  |              | | ||||   ||:
Db     848 HNAESSASGIPLVQARSNTVNGGAPVPPAPGS-----------GATGSGTSGSGTSESVT 896

Qy      62 KMAAAWPSGPS-----------------APEAVTARLVGVLWFVSVTTGPWGAVATSAG 103
             : |  || :                  |  | |: : |     :|     ||   : |
Db     897 NGSGATESGSTGSGTTGTGTSGTGSSGTGASAARTSSIAGDAPQAAVLADTPGAAGAGG 956

Qy     104 GE------ESL--------KCEDLKVGQYICKDPKINDAT---------QEPVNCTNYTA 140
           |       :||         : ::|::| |:       |             :||   ||:
Db     957 GRSNCFSADSLVTTVTGQKRMDELQIGDYVLVPSSGNVLKYEKVEMFYHREPKTRTNF-- 1015

Qy     141 HVSCFPAPNITCKDSSGNETHFTGNEVGFFKPIS-CRNVNGYS 182
                 :     || :   ||  :    |:: |  |  |:
Db    1015 ---------VVLYTKSGRKLSLTGRHL---LPVAECSQVEQYT 1045


RESULT   7
Q9Z337
ID   Q9Z337      PRELIMINARY;      PRT;  132 AA.
AC   Q9Z337;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   LECT2Q PROTEIN.
GN   LECT2Q.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129; TISSUE=LIVER;
RX   MEDLINE; 98382586.
RA   YAMAGOE S., WATANABE T., MIZUNO S., SUZUKI K.;
RT   "The mouse Lect2 gene: cloning of cDNA and genomic DNA, structural
RT   characterization and chromosomal localization.";
RL   Gene 216:171-178(1998).
DR   EMBL; AB009689; BAA33386.1; -.
SQ   SEQUENCE   132 AA;  14601 MW;  AD0C6569 CRC32;


  Query Match             6.0%;  Score 86;  DB 11;  Length 132;
  Best Local Similarity   27.0%;  Pred. No. 1.2;
  Matches   27;  Conservative    7;  Mismatches   48;  Indels   18;  Gaps    3;

Qy      78 TARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCTN 137
           |  |:      |   |||  :  |    |   |:    |||  :       ||
Db       4 TTILISAALLSSALAGPWANICASKSSNEIRTCDSYGCGQYSAQ------RTQR------ 52

Qy     138 YTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRN 177
             |    |  :: | | |     |||  ||  ||   :|
Db      52 ---H---HPGVDVLCSDGSVVYAPFTGKIVGQEKPYRNKN 85


RESULT   8
Q9ZBV5
ID   Q9ZBV5      PRELIMINARY;      PRT;  446 AA.
AC   Q9ZBV5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE INTEGRAL MEMBRANE TRANSPORTER.
GN   SC4B5.13.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   SEEGER K.J., HARRIS D.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
```

```
ID   Q66668      PRELIMINARY;      PRT;   456 AA.
AC   Q66668;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ORF 66.
OS   Equine herpesvirus 2.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=86/67;
RX   MEDLINE; 95302501.
RA   TELFORD E.A., WATSON M.S., AIRD H.C., PERRY J., DAVISON A.J.;
RT   "The DNA sequence of equine herpesvirus 2.";
RL   J. Mol. Biol. 249:520-528(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=86/67;
RA   TELFORD E.A.R.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U20824; AAC13854.1; -.
SQ   SEQUENCE   456 AA;  49776 MW;  E140A63B CRC32;

  Query Match             6.4%;  Score 92;  DB 12;  Length 456;
  Best Local Similarity   26.2%;  Pred. No. 1.5;
  Matches   42;  Conservative   11;  Mismatches   45;  Indels   62;  Gaps    7;

Qy      47 LGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLWFVSVTTGPWGAVATSAGGEE 106
           ||||| | |:                        |:|                   |||
Db      13 LGGGGGGGGD-----------------------LLG------------------GGEA 29

Qy     107 S----LKCEDLKVGQYIC-KDPKINDATQEPVNCTNYTAHV------SCFPAPNITCKDS 155
               || |:||:  | :      |  :   |  : ||:      |||    :||  |
Db      30 DGLMRALCEGLRVGEEDCARFVLYGVAYWQGGRCPEWVAHITRCADLSCFATYLLTCHRS 89

Qy     156 SGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGW 195
           | |  |||  |   :  | |        | |  |||| :
Db      90 GGCE--FTGGRVARDRLPSLRE------SVEVLQSLFLAF 121


RESULT   3
O22512
ID   O22512      PRELIMINARY;      PRT;   690 AA.
AC   O22512;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GRR1.
GN   GRR1.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Glycine.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ROOT;
RA   CHEN W., ATHERLY A.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF019910; AAB70660.1; -.
SQ   SEQUENCE   690 AA;  73766 MW;  DB7DD87A CRC32;

  Query Match             6.2%;  Score 89.5;  DB 10;  Length 690;
  Best Local Similarity   24.6%;  Pred. No. 4.1;
  Matches   43;  Conservative   22;  Mismatches   59;  Indels   51;  Gaps    7;

Qy      85 LWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPV--NCTNYTAHV 142
           || |: | |  | :  ::|  :  | :       :|| | |:| |   |  || |  | :
Db     233 LWDVA-TVGDVGLIEIASGCHQLEKLD-------LCKCPNISDKTLIAVAKNCPN-LAEL 283

Qy     143 SCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFY 202
           |    |||      |||      :    : || :| :|   :
Db     284 SIESCPNI------GNEGLQAIGKCPNLRSISIKNCSGVGDQ----------------- 320

Qy     203 LGYPALGLLKFCTVGFCGIGSLIDFIL--ISMQIVGPSDGSSYIIDYYGTRLTRL 255
                       |  |: |   | |  : :: :  || |  :| :||  :| |
Db     320 -------------GVAGLLSSASFALTKVKLESLTVSDLSLAVIGHYGVAVTDL 360


RESULT   4
Q98339
ID   Q98339      PRELIMINARY;      PRT;   330 AA.
AC   Q98339;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   SPIKE PROTEIN (FRAGMENT).
OS   murine hepatitis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHV-3;
RX   MEDLINE; 96183918.
RA   SUZUKI H., TAGUCHI F.;
RT   "Analysis of the receptor-binding site of murine coronavirus spike
RT   protein.";
RL   J. Virol. 70:2632-2636(1996).
DR   EMBL; D83335; BAA11891.1; -.
FT   NON_TER     330     330
SQ   SEQUENCE   330 AA;  36741 MW;  BB58D034 CRC32;

  Query Match             6.2%;  Score 89;  DB 12;  Length 330;
  Best Local Similarity   23.6%;  Pred. No. 1.9;
  Matches   59;  Conservative   27;  Mismatches   94;  Indels   70;  Gaps   13;

Qy      25 TGLYPMRG-PFKNLALLP--------FSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSA-- 74
           || ||: |  |:||||          |  | |      |   |  |  :  : ||| :|
Db      66 TGYYPVDGSKFRNLALTGTNSVSLSWFQPPYLSQFNDGIFAK--VQNLKTSTPSGATAYF 123

Qy      74 PEAVTARLVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQY-ICKDPKINDATQEP 132
           |  |   | |   : :|   |:  |  ::            | || ||: |
Db     124 PTIVIGSLFGYTSY-TVVIEPYNGVIMAS-----------VCQYTICQLP--------- 162

Qy     133 VNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE-VGFF----KPISCRNVNGYSYKVAV 187
                :  | |  |                  ||: :||:    ||  |    ::  |
Db     162 --------YTDCKPNTN-------------GNKLIGFWHTDVKPPICVLKRNFTLNVNA 199

Qy     188 ALSLFLGWLGADRFYLGY---PALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
               |  :     ||  |   |:     |       || ::    :   |  |: ||::
Db     200 DAFYFHFYQHGGTFYAYYADKPSATTFLFSVY----IGDILTQYYVLPFICNPTAGSTFR 255

Qy     245 IDYYGTRLTR 254
           |: | | :
Db     256 PRYWVTPLVK 265


RESULT   5
O39227
ID   O39227      PRELIMINARY;      PRT;  1324 AA.
AC   O39227;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   E2 GLYCOPROTEIN PRECURSOR.
OS   murine hepatitis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHV-A59;
RA   LEPARC-GOFFART I., HINGLEY S.T., CHUA M.M., JIANG X., LAVI E.,
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: February 11, 2000, 04:33:16 ; Search time 14.98 Seconds
(without alignments)
1245.052 Million cell updates/sec

Title: US-09-060-609-2
Perfect score: 1439
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 225878 seqs, 69334122 residues

Database : SPTREMBL_12:*

Word size : 0

Number of hits that pass the threshold : 225878
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	95.5	6.6	913	4	O75061	O75061 homo sapien
2	92	6.4	456	12	Q66668	Q66668 equine herp
3	89.5	6.2	690	10	O22512	O22512 glycine max
4	89	6.2	330	12	Q98339	Q98339 murine hepa
5	88	6.1	1324	12	O39227	O39227 murine hepa
6	87.5	6.1	1207	5	Q21535	Q21535 caenorhabdi
7	86	6.0	132	11	Q9Z337	Q9z337 mus musculu
8	84	5.8	446	2	Q9ZBV5	Q9zbv5 streptomyce
9	83	5.8	1274	10	P93125	P93125 dunaliella
10	82.5	5.7	789	2	Q9XBP8	Q9xbp8 myxococcus
11	82	5.7	409	10	P93368	P93368 nicotiana t
12	82	5.7	465	11	O88354	O88354 spermophilu
13	81.5	5.7	644	12	P87656	P87656 eggdrop syn
14	81.5	5.7	644	12	O11424	O11424 avian adeno
15	81.5	5.7	3432	12	O90297	O90297 japanese en
16	80.5	5.6	455	4	Q99808	Q99808 homo sapien
17	80.5	5.6	377	5	Q9XY68	Q9xy68 balanus bal
18	80.5	5.6	3432	12	O90417	O90417 japanese en
19	80.5	5.6	1237	13	Q91976	Q91976 gallus gall
20	80	5.6	638	5	O77408	O77408 lymnaea sta
21	80	5.6	171	10	O81433	O81433 arabidopsis
22	79.5	5.5	489	2	O33261	O33261 mycobacteri
23	79.5	5.5	171	9	Q38215	Q38215 insertion s
24	79.5	5.5	530	12	Q83876	Q83876 norwalk vir
25	79.5	5.5	3432	12	Q82873	Q82873 japanese en
26	79.5	5.5	3432	12	Q82872	Q82872 japanese en
27	79.5	5.5	3432	12	Q9WCX7	Q9wcx7 japanese en
28	79.5	5.5	1064	12	Q9WAU6	Q9wau6 broad bean
29	79	5.5	1278	4	O60434	O60434 homo sapien
30	79	5.5	797	4	O95288	O95288 homo sapien
31	79	5.5	3507	5	Q23587	Q23587 caenorhabdi
32	78.5	5.5	482	2	Q9ZMT7	Q9zmt7 helicobacte
33	78.5	5.5	7962	4	Q10465	Q10465 homo sapien
34	78.5	5.5	5376	11	O88799	O88799 mus musculu
35	78	5.4	419	2	P72207	P72207 pseudomonas
36	78	5.4	523	10	Q39438	Q39438 beta vulgar
37	78	5.4	3019	12	O92529	O92529 hepatitis c
38	78	5.4	414	12	Q81267	Q81267 hepatitis c
39	78	5.4	446	13	Q9YIB2	Q9yib2 anguilla ja
40	77.5	5.4	374	1	O26843	O26843 methanobact
41	77.5	5.4	869	5	Q9Y0Z8	Q9y0z8 drosophila
42	77	5.4	797	4	Q9Y345	Q9y345 homo sapien
43	77	5.4	1239	5	O61541	O61541 drosophila
44	77	5.4	1302	5	O61542	O61542 drosophila
45	77	5.4	1049	10	Q9ZNR9	Q9znr9 volvox cart

ALIGNMENTS

RESULT 1
O75061
ID O75061 PRELIMINARY; PRT; 913 AA.
AC O75061;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE KIAA0473 PROTEIN.
GN KIAA0473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98116662.
RA SEKI N., OHIRA M., NAGASE T., ISHIKAWA K., MIYAJIMA N., NAKAJIMA D.,
RA NOMURA N., OHARA O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
DR EMBL; AB007942; BAA32318.1; -.
SQ SEQUENCE 913 AA; 99996 MW; 025A0952 CRC32;

Query Match 6.6%; Score 95.5; DB 4; Length 913;
Best Local Similarity 24.9%; Pred. No. 1.7;
Matches 48; Conservative 23; Mismatches 87; Indels 35; Gaps 8;

```
Qy     29 PMRGPFKNLALLPFSLPLLGGGGSGSGEKVSVSKMAAAW-PSGPSAPEAVTAR------- 81
          |   |  |  ||     | ||||:    :   | :   : |||||:: ::   |
Db    513 PPAAPPTNSELLS---DLFGGGGAAGPTQAGQSGVEDVFHPSGPASTQSTPRRSATSTSA 569

Qy     81 ----LVGVLWFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCT 136
              ||       |  |:|| :  :|       :|| :| ::      :|   :|
Db    570 SPTLRVG----EGATFDPFGAPSKPSG-------QDL-LGSFLNTSSASSDPFLQPTRSP 617

Qy    137 NYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAV------ALS 190
          : | | |  || ||    | | : |      |     :  :  | |:           |
Db    618 SPTVHASSTPAVNIQPDVSGGWDWHAKPGGPGMGSKSAATSPTGSSHGTPTHQSKPQTLD 677

Qy    191 LF--LGWLGADRF 201
            |  || ||:  |
Db    678 PFADLGTLGSSSF 690
```

RESULT 2
Q66668

Search completed: February 11, 2000, 03:33:27
Job time: 3116 sec

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1682498"
                     /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      111 a    110 c    120 g    130 t
ORIGIN

Query Match             54.0%;  Score 437.8;  DB 43;  Length 471;
  Best Local Similarity   98.4%;  Pred. No. 6.5e-108;
  Matches  442;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTT 638
          |||||||||||||||| ||  |  |  ||
Db    421 GGATACCCTGCTTTGGATTGTTGGACCTT 449


RESULT  15
AI299329
LOCUS       AI299329      423 bp    mRNA            EST       01-FEB-1999
DEFINITION  qn22c06.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1898986 3',
            mRNA sequence.
ACCESSION   AI299329
VERSION     AI299329.1  GI:3958983
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1021   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1898986"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      101 a     98 c    107 g    117 t
ORIGIN


  Query Match             52.2%;  Score 423;  DB 44;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-104;
  Matches  423;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    208 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 60

Qy    268 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 120

Qy    328 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAAATAAATGACGCTACG 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAAATAAATGACGCTACG 180

Qy    388 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 240

Qy    448 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

Qy    508 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 360

Qy    568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 420

Qy    628 TTG 630
          |||
Db    421 TTG 423
```

```
Qy    208 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 60

Qy    268 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 120

Qy    328 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 180

Qy    388 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 240

Qy    448 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

Qy    508 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 360

Qy    568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 420

Qy    628 TTGTTAAAGTTTTGCACTG 646
          ||||||||||||||||||
Db    421 TTGTTAAAGTTTTGCACTG 439
```

RESULT 13
AI038331
LOCUS AI038331 448 bp mRNA EST 24-SEP-1998
DEFINITION ox84a06.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:1662994 3', mRNA sequence.
ACCESSION AI038331
VERSION AI038331.1 GI:3277525
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 448)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802984.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1099 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 432.
FEATURES Location/Qualifiers
source 1. .448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1662994"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 118 a 92 c 105 g 133 t
ORIGIN

Query Match 54.1%; Score 438.2; DB 41; Length 448;
Best Local Similarity 99.3%; Pred. No. 5e-108;
Matches 440; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy    258 GTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGA 317
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGA 60

Qy    318 GTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAA 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAA 120

Qy    378 TGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCC 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCC 180

Qy    438 AGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGA 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGA 240

Qy    498 AGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGC 557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGC 300

Qy    558 AGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCC 617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCC 360

Qy    618 TGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGA 677
          ||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db    361 TGCTTTGGGTTTGTTAAAGTTTTGCACTGTATGGTTATGTGGAATTGGGAGCCTAATTGA 420

Qy    678 TTTCATTCTTATTTCAATGCAGA 700
          ||||||||||||||||||||||
Db    421 TTTCATTCTTATTTCAATGCAGA 443
```

RESULT 14
AI168073
LOCUS AI168073 471 bp mRNA EST 05-OCT-1998
DEFINITION oz89c02.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:1682498 3', mRNA sequence.
ACCESSION AI168073
VERSION AI168073.1 GI:3701243
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 471)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2282128.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES Location/Qualifiers
source 1. .471

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., ,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1288147.
            Other_ESTs: THC195202
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .475
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):108677"
                      /db_xref="taxon:9606"
                      /clone_lib="Colon carcinoma (HCC) cell line"
                      /tissue_type="colon"
                      /cell_type="KM12C"
                      /cell_line="KM12C(HCC)-parental human colon
                      carcinoma;Dukes B2"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
BASE COUNT      108 a    107 c    119 g    135 t      6 others
ORIGIN
```

```
  Query Match             54.9%;  Score 444.4;  DB 31;  Length 475;
  Best Local Similarity   98.1%;  Pred. No. 1.1e-109;
  Matches  467;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy    189 GGCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGG 248
          ||||||||||||||  |||||||||||||||||||||||||||||| ||||||||||||
Db      1 GGCGGCCGCCTGGCCTNCTGGTCCGTCTGCTCCGGAGGCCGTGACGG-CAGACTCGTTGG 59

Qy    249 TGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGG 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGG 119

Qy    309 GGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCC 368
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    120 GGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTNTAAAGATCC 179

Qy    369 AAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTC 428
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 AAAAATAAATNACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTC 239

Qy    429 CTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTAC 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 CTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTAC 299

Qy    489 TGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTA 548
          |||||||||| | | ||||||||||||||||||||||||||||||||||||||||||||
Db    300 TGGGAACGAANTNGNTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTA 359

Qy    549 CAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCT 608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 CAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCT 419

Qy    609 TGGATACCCTGCTTTGGGTTTGTTAAAG-TTTTGCACTGTAGGGTTTTGTGGAATT 663
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    420 TGGATACCCTGCTTTGGGTTTGTTAAAGTTTTTGCACTGTAGGGTTTTGTGGAATT 475
```

```
RESULT  12
AI299154
LOCUS       AI299154      439 bp    mRNA            EST       29-JAN-1999
DEFINITION  qm96f11.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1896621 3',
            mRNA sequence.
ACCESSION   AI299154
VERSION     AI299154.1  GI:3958808
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 966   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 419.
FEATURES              Location/Qualifiers
     source           1. .439
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1896621"
                      /clone_lib="NCI_CGAP_Lu5"
                      /tissue_type="carcinoid"
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; 1st strand cDNA was prepared from
                      neuroendocrine lung carcinoid, and was then primed with a
                      Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                      to Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector.  Library is normalized. Library was
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      105 a    100 c    110 g    124 t
ORIGIN
```

```
  Query Match             54.2%;  Score 439;  DB 44;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 3e-108;
  Matches  439;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCTTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGNGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
          |||||||||||||||||||||||||||||||||  | ||||||||||||||||||||||
Db    421 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTANNGGTTTGTGGAATTGGGAGC 480

Qy    670 CTAATTGATTTCATTCTTAT 689
          |||||||| ||||||||||||
Db    481 CTAATTGANTTCATTCTTAT 500


RESULT  10
N47594
LOCUS       N47594        507 bp    mRNA            EST       14-FEB-1996
DEFINITION  yy59a07.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA
            clone IMAGE:277812 3', mRNA sequence.
ACCESSION   N47594
VERSION     N47594.1  GI:1188760
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:802394.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyT not found
            Seq primer: m13 -40 forward
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /db_xref="GDB:3896188"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:277812"
                     /clone_lib="Soares_multiple_sclerosis_2NbHMSP"
                     /sex="male"
                     /tissue_type="multiple sclerosis lesions"
                     /dev_stage="Age 46"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                     lesions from one patient was kindly provided by Dr. Kevin
                     G. Becker (NINDS/NIH). "
BASE COUNT      117 a    109 c    124 g    148 t      9 others
ORIGIN


  Query Match          56.0%;  Score 453.2;  DB 25;  Length 507;
  Best Local Similarity   95.3%;  Pred. No. 4.6e-112;
  Matches  484;  Conservative    0;  Mismatches   21;  Indels    3;  Gaps    2;

Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||| ||||||||      ||||||||||||||||||  |      ||   ||||||||||
Db      1 GCGNCCGCCTGGNACNTNGGTCCGTCTGCTCCGGAG--CCGTGACGCAGNACTCGTTGGT 58

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 118

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          |     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 GNNAGNGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 178

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 238

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 298

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 358

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 418

Qy    610 GGATACCCTGCTTT-GGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAG 668
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    419 GGATACCCTGCTTTGGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAG 478

Qy    669 CCTAATTGATTTCATTCTTATTTCAATG 696
          ||||||||||||||||||||||||||||
Db    479 CCTAATTGATTTCATTCTTATTTCAATG 506


RESULT  11
AA306979
LOCUS       AA306979      475 bp    mRNA            EST       18-APR-1997
DEFINITION  EST178050 Colon carcinoma (HCC) cell line Homo sapiens cDNA 5' end,
            mRNA sequence.
ACCESSION   AA306979
VERSION     AA306979.1  GI:1959469
KEYWORDS    EST.
SOURCE      human.
```

```
AI349520
LOCUS       AI349520      498 bp    mRNA            EST       03-FEB-1999
DEFINITION  qp72g05.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1928600 3', mRNA sequence.
ACCESSION   AI349520
VERSION     AI349520.1  GI:4086726
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1009   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1928600"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT      113 a    112 c    128 g    144 t      1 others
ORIGIN


  Query Match             61.4%;  Score 497;  DB 45;  Length 498;
  Best Local Similarity   99.8%;  Pred. No. 6.7e-124;
  Matches  497;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy     250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy     310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy     370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy     430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy     490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy     550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy     610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGATACCCTGCTNTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 480

Qy     670 CTAATTGATTTCATTCTT 687
           ||||||||||||||||||
Db     481 CTAATTGATTTCATTCTT 498


RESULT   9
AI680969
LOCUS       AI680969      500 bp    mRNA            EST       26-MAY-1999
DEFINITION  tx35h07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2271613 3',
            mRNA sequence.
ACCESSION   AI680969
VERSION     AI680969.1  GI:4891151
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135321.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2271613"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      114 a    112 c    125 g    145 t      4 others
ORIGIN


  Query Match             60.8%;  Score 492.8;  DB 50;  Length 500;
  Best Local Similarity   98.8%;  Pred. No. 9.2e-123;
  Matches  494;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
```

cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 146 a 106 c 126 g 169 t 1 others
ORIGIN

Query Match 63.4%; Score 513.4; DB 38; Length 548;
Best Local Similarity 98.2%; Pred. No. 2.6e-128;
Matches 540; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

```
Qy    240 ACTCGTTGGTGTCCTGTGGTTCGTCT-CAGTCACTACAGGACCCTGGGGGGCTGTTGCCA 298
          ||||||||||||||||||||||||||| |||||||||||||||||||||||    |   ||
Db      1 ACTCGTTGGTGTCCTGTGGTTCGTCTGCAGTCACTACAGGACCCTGGGGGCTGTTG--CA 58

Qy    299 CCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTT 358
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db     59 CCTCCGCCGGGGGCANGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTT 118

Qy    359 GTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAG 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 GTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAG 178

Qy    419 CTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAA 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 CTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAA 238

Qy    479 CACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATG 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 CACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATG 298

Qy    539 GCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATC 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 GCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATC 358

Qy    599 GATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTG 658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 GATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTG 418

Qy    659 GAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATG 718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 GAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATG 478

Qy    719 GAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATG 778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 GAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATG 538

Qy    779 AAACATTTAG 788
          ||||||||||
Db    539 AAACATTTAG 548
```

RESULT 7
AI682204
LOCUS AI682204 507 bp mRNA EST 26-MAY-1999
DEFINITION wa71b06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2301587 3', mRNA sequence.
ACCESSION AI682204
VERSION AI682204.1 GI:4892386
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT On Dec 20, 1995 this sequence version replaced gi:1136077.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 469.
FEATURES Location/Qualifiers
source 1. .507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2301587"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 120 a 109 c 125 g 152 t 1 others
ORIGIN

Query Match 62.5%; Score 506; DB 50; Length 507;
Best Local Similarity 99.8%; Pred. No. 2.5e-126;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    208 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCA 60

Qy    268 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAG 120

Qy    328 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACG 180

Qy    388 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAAC 240

Qy    448 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTT 300

Qy    508 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTG 360

Qy    568 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGT 420

Qy    628 TTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 687
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    421 TTGTTAAAGTTTTGCACTGTANGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTT 480

Qy    688 ATTTCAATGCAGATTGTTGGACCTTCA 714
          |||||||||||||||||||||||||||
Db    481 ATTTCAATGCAGATTGTTGGACCTTCA 507
```

RESULT 8

```
AI674462
LOCUS       AI674462      530 bp    mRNA            EST       19-MAY-1999
DEFINITION  wc44e01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321496 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI674462
VERSION     AI674462.1  GI:4874942
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948555.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2321496"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5;000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      123 a    116 c    133 g    157 t      1 others
ORIGIN
```

Query Match 65.1%; Score 527.4; DB 50; Length 530;
Best Local Similarity 99.6%; Pred. No. 4.2e-132;
Matches 528; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    194 CCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCC 253
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      1 CCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTTACGGCCAGACTCGTTGGTGTCC 60

Qy    254 TGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCG 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCG 120

Qy    314 AGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAA 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAA 180

Qy    374 TAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTT 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTT 240

Qy    434 TTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGA 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGA 300

Qy    494 ACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAG 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAG 360

Qy    554 TGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGAT 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGAT 420

Qy    614 ACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAA 673
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    421 ACCCTGCTTTGGGTTTGTTAAAGNTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAA 480

Qy    674 TTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGT 723
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGT 530
```

```
RESULT 6
AA772225
LOCUS       AA772225      548 bp    mRNA            EST       31-DEC-1998
DEFINITION  ai41c01.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            1359552 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;,
            mRNA sequence.
ACCESSION   AA772225
VERSION     AA772225.1  GI:2824008
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811076.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Insert Length: 978   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="1359552"
                     /clone_lib="Soares_parathyroid_tumor_NbHPA"
                     /tissue_type="parathyroid tumor"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: parathyroid gland; Vector: pT7T3D
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer
                     [5'-
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTTT
                     T-3'], double-stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
```

Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 148 a 109 c 130 g 168 t 1 others
ORIGIN

Query Match 68.5%; Score 555; DB 47; Length 556;
Best Local Similarity 99.8%; Pred. No. 1.5e-139;
Matches 555; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    234 GGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGT 293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGT 60

Qy    294 TGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATA 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATA 120

Qy    354 TATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTA 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TATTTGTAAAGATCCAAAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTA 180

Qy    414 CACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAA 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAA 240

Qy    474 TGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGT 533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGT 300

Qy    534 AAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGC 593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGC 360

Qy    594 AGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTT 653
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGATCGATTNTACCTTGGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTT 420

Qy    654 TTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTC 713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTC 480

Qy    714 AGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTAC 773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGATGGAAGTAGTTACATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTAC 540

Qy    774 TAATGAAACATTTAGA 789
          ||||||||||||||||
Db    541 TAATGAAACATTTAGA 556
```

RESULT 4
AI580361
LOCUS AI580361 537 bp mRNA EST 05-APR-1999
DEFINITION tm47h02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2161299 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION AI580361
VERSION AI580361.1 GI:4564737
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 537)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949103.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 492.
FEATURES Location/Qualifiers
source 1. .537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2161299"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 124 a 117 c 139 g 156 t 1 others
ORIGIN

Query Match 65.6%; Score 531.2; DB 48; Length 537;
Best Local Similarity 99.3%; Pred. No. 4e-133;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGATACCCTGCTNTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 480

Qy    670 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGT 726
          ||||||||||||||||||||||||||||||||||||||||||||||||| |  |||
Db    481 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGGAAGAGT 537
```

RESULT 5

```
LOCUS       AI680904      574 bp    mRNA            EST       26-MAY-1999
DEFINITION  tx42f05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272257 3'
            similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION   AI680904
VERSION     AI680904.1  GI:4891086
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1134035.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2272257"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      148 a    118 c    136 g    172 t
ORIGIN
```

Query Match 70.7%; Score 572.4; DB 50; Length 574;
Best Local Similarity 99.8%; Pred. No. 3e-144;
Matches 573; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    211 CCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTC 270
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGTCTGCTCCGGAGGCCGTGACTGCCAGACTCGTTGGTGTCCTGTGGTTCGTCTCAGTC 60

Qy    271 ACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGC 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGGGGCGAGGAGTCGCTTAAGTGC 120

Qy    331 GAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAA 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCAAAAATAAATGACGCTACGCAA 180

Qy    391 GAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATA 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCCTGTTTTCCAGCACCCAACATA 240

Qy    451 ACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTC 510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACTGGGAACGAAGTTGGTTTTTTC 300

Qy    511 AAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCT 570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTACAAAGTGGCAGTCGCATTGTCT 360

Qy    571 CTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTG 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCTGCTTTGGGTTTG 420

Qy    631 TTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATT 690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGCCTAATTGATTTCATTCTTATT 480

Qy    691 TCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACC 750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGATTACTATGGAACC 540

Qy    751 AGACTTACAAGACTGAGTATTACTAATGAAACAT 784
          ||||||||||||||||||||||||||||||||||
Db    541 AGACTTACAAGACTGAGTATTACTAATGAAACAT 574
```

```
RESULT   3
AI479764
LOCUS       AI479764      556 bp    mRNA            EST       14-APR-1999
DEFINITION  tm69b04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163343 3',
            mRNA sequence.
ACCESSION   AI479764
VERSION     AI479764.1  GI:4372932
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188819.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 614   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2163343"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
```

	8	497	61.4	498	45	AI349520 AI349520 qp72g05.x
	9	492.8	60.8	500	50	AI680969 AI680969 tx35h07.x
	10	453.2	56.0	507	25	N47594 N47594 yy59a07.s1
	11	444.4	54.9	475	31	AA306979 AA306979 EST178050
	12	439	54.2	439	44	AI299154 AI299154 qm96f11.x
	13	438.2	54.1	448	41	AI038331 AI038331 ox84a06.x
	14	437.8	54.0	471	43	AI168073 AI168073 oz89c02.x
	15	423	52.2	423	44	AI299329 AI299329 qn22c06.x
	16	396.6	49.0	674	29	AA143062 AA143062 zl49h02.s
	17	395	48.8	395	44	AI245878 AI245878 qk27f08.x
	18	388	47.9	508	36	AA639448 AA639448 nq88g05.s
	19	387.6	47.9	540	44	AI284315 AI284315 qj65c02.x
	20	385	47.5	398	64	AW071521 AW071521 xa36a07.x
c	21	384.4	47.5	488	27	AA017855 AA017855 mh47c09.r
c	22	382.6	47.2	476	27	AA020016 AA020016 mh48a12.r
	23	379	46.8	416	38	AA757905 AA757905 zg42h09.s
	24	377.8	46.6	474	43	AI239974 AI239974 qh44d02.x
	25	368	45.4	461	50	AI693751 AI693751 wd44h04.x
	26	366	45.2	385	25	D81362 D81362 HUM159G08B
	27	356.8	44.0	502	43	AI175138 AI175138 EST218660
	28	355.2	43.9	417	23	D61247 D61247 HUM172D06B
	29	348	43.0	639	30	AA218227 AA218227 mv61g08.r
	30	346.6	42.8	464	24	N34957 N34957 yy50b11.s1
	31	340.8	42.1	382	32	AA345598 AA345598 EST51648
	32	331.8	41.0	392	38	AA757694 AA757694 zg38h03.s
	33	324	40.0	335	43	AI194064 AI194064 qe78d08.x
	34	322	39.8	323	62	AI903940 AI903940 MR-BT038-
	35	321.6	39.7	457	40	AA943663 AA943663 EST199162
	36	320.4	39.6	322	62	AI903935 AI903935 MR-BT038-
	37	320.4	39.6	323	62	AI903938 AI903938 MR-BT038-
	38	319.2	39.4	580	23	R99199 R99199 yq64d10.r1
c	39	306	37.8	632	41	AI057115 AI057115 oz23g10.x
	40	294	36.3	420	37	AA711942 AA711942 vu59h08.r
c	41	292.8	36.1	356	27	W84976 W84976 mf42g02.r1
	42	288.4	35.6	418	43	AI176189 AI176189 EST219770
	43	288	35.6	300	62	AI887092 AI887092 wl96f10.x
c	44	265.2	32.7	342	27	AA023218 AA023218 mh69b08.r
	45	262.8	32.4	407	23	H52616 H52616 yq83e09.r1

ALIGNMENTS

RESULT 1
AI923178
LOCUS AI923178 599 bp mRNA EST 02-SEP-1999
DEFINITION wn67b10.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2450491 3' similar to WP:C02F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.
ACCESSION AI923178
VERSION AI923178.1 GI:5659142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 599)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133491.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 457.

FEATURES Location/Qualifiers
source 1. .599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2450491"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went througb one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 148 a 127 c 146 g 176 t 2 others
ORIGIN

Query Match 73.5%; Score 595.4; DB 62; Length 599;
Best Local Similarity 99.5%; Pred. No. 1.8e-150;
Matches 596; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy    190 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCGCCTGGCCGTCTGGTCCGTCTGCTCCGGAGGCCGTGACGGCCAGACTCGTTGGT 60

Qy    250 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCTGTGGTTCGTCTCAGTCACTACAGGACCCTGGGGGGCTGTTGCCACCTCCGCCGGG 120

Qy    310 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGAGGAGTCGCTTAAGTGCGAGGACCTCAAAGTGGGACAATATATTTGTAAAGATCCA 180

Qy    370 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAATAAATGACGCTACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTTCC 240

Qy    430 TGTTTTCCAGCACCCAACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 489
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTTTTCCAGCACCCCACATAACTTGTAAGGATTCCAGTGGCAATGAAACACATTTTACT 300

Qy    490 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGAACGAAGTTGGTTTTTTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCCTAC 360

Qy    550 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAGTGGCAGTCGCATTGTCTCTTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 420

Qy    610 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTAGGGTTTTGTGGAATTGGGAGC 669
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    421 GGATACCCTGCTTTGGGTTTGTTAAAGTTTTGCACTGTANGGTTTTGTGGAATTGGGAGC 480

Qy    670 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTAC 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGTTAC 540

Qy    730 ATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACATTTAG 788
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    541 ATTATAGATTACTATGGAACCAGACTTACAAGACTGAGTATTACTAATGAAACATNTAG 599
```

RESULT 2
AI680904

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 02:41:31 ; Search time 506.49 Seconds
(without alignments)
6038.200 Million cell updates/sec

Title: US-09-060-609-1
Perfect score: 810
Sequence: 1 ATGCATATTTTAAAAGGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	595.4	73.5	599	62	AI923178	AI923178 wn67b10.x
2	572.4	70.7	574	50	AI680904	AI680904 tx42f05.x
3	555	68.5	556	47	AI479764	AI479764 tm69b04.x
4	531.2	65.6	537	48	AI580361	AI580361 tm47h02.x
5	527.4	65.1	530	50	AI674462	AI674462 wc44e01.x
6	513.4	63.4	548	38	AA772225	AA772225 ai41c01.s
7	506	62.5	507	50	AI682204	AI682204 wa71b06.x